Study of the Associated Production of a Z Boson with b Hadrons in Proton-Proton Collisions with the CMS Experiment

Dissertation

zur

Erlangung der naturwissenschaftlichen Doktorwürde
(Dr. sc. nat.)

vorgelegt der

Mathematisch-naturwissenschaftlichen Fakultät

der

Universität Zürich

von

Carlotta Favaro
aus

Italien

Promotionkomitee

Prof. Dr. V. Chiochia
Dr. S. de Visscher

Zürich, 2013

Contents

ABSTRACT

Diese Dissertation präsentiert die Messungen von Winkel-Korrelationen und des totalen Wirkungsquerschnitts des pp $\rightarrow$ Zb$\bar{\text{b}}$X Prozesses bei einer Schwerpunktsenergie von $\sqrt{s} =$ 7 TeV.

Die assoziierte Produktion von Vektor-Bosonen und schweren Quarks ist einer der wichtigsten Untergrund-Prozesse für die Entdeckung und Untersuchung des Higgs Bosons im Standard Modell, beim Zerfall in ein b$\bar{\text{b}}$ Quark Paar. Nichtsdestotrotz ist der Prozess bisher wenig verstanden, und die Modellierung der Produktion von kolinearen b-Quark Paaren ist beeinträchtigt von erheblichen theoretischen Unsicherheiten. Mehrere unterschiedliche Berechnungs-Schemen werden in Monte Carlo Generatoren angewandt um die Zb$\bar{\text{b}}$ Signatur am Large Hadron Collider zu simulieren, und durch die Untersuchung der
Winkel-Korrelationen kann die Simulations-Technik mit der besten Beschreibung der Daten ermittelt werden.

Die Messung wird durchgeführt mit einem Datensatz des CMS Experiments am LHC der 5.15 fb^{-1} entspricht. Die inklusiven und differentiellen Wirkungsquerschnitte werden in verschiedenen kinematischen Bereichen gemessen, welche durch eine Selektion des transversalen Impulses des Z Bosons festgelegt werden.

Um Kollisionen mit einem Z Boson-Zerfall zu selektieren wird die Präsenz von zwei Leptonen, Müonen oder Elektronen, verlangt. Hadronen die ein b Valenz-Quark enthalten, werden mit einer Technik identifiziert die, da sie von der Jet Rekonstruktion völlig unabhängig ist, zum ersten Mal die Sensitivität bis zum Bereich der kolinearen Produktion von b-Hadron Paaren erweitert.

Die Messungen der Wirkungsquerschnitte werden mit den theoretischen Vorhersagen von mehreren Monte Carlo Generatoren verglichen, welche unterschiedliche Berechnungs-Schemen implementieren, in führender und nachführender Ordnung der störungs-theoretischen Erweiterung: MadGraph, Alpgen und aMC@NLO. Die gemessenen Winkelverteilungen sind in innerhalb der experimentellen und theoretischen Unsicherheiten in guter Übereinstimmung mit den Erwartungen. Die beste Beschreibung der Daten liefert der Alpgen Generator.
Lokalisierte Diskrepanzen werden dennoch gesehen, speziell in der Produktion von kolinearen

b-Hadron Paaren.

Die Beobachtungen sind konsistent mit der Hypothese einer unzureichenden Modellierung der Kontributionen von $qq \to Zb\bar{b}X$ und $qg \to Zb\bar{b}X$ Unterprozessen, welche ein Feynman Diagramm beinhalten mit einem Vertex wo ein sich ein Gluon in zwei b-Quarks aufteilt, $g \to b\bar{b}$. Überdies wird eine Diskrepanz des totalen Wirkungsquerschnitt beobachtet, in dem die Daten alle theoretischen Vorhersagen um etwa 15% übersteigen. Hingegen stimmt der Verlauf des Wirkungsquerschnitts als Funktion des Transversal-Impulses des Z Bosons mit den Erwartungen überein.

Diese Arbeit präsentiert ausserdem eine Messung der räumlichen Auflösung welche vom Barrel Teil des Pixel-Detektors von CMS erreicht wurde, durchgeführt mit LHC Kollisionsdaten. Die Auflösung der Pixel Hits beinträchtigt massgeblich die Leistung der Spuren- und Vertex-Rekonstruktion, welche wiederum entscheidend ist für die zuverlässige Identifikation von Zerfällen von schweren Quarks.

ABSTRACT

This thesis presents the measurement of the angular correlations and total cross section of the process $pp \to Zb\bar{b}X$ at a center-of-mass energy $\sqrt{s} = 7$ TeV.

The associated production of vector bosons and heavy quarks constitutes one of the main backgrounds for the discovery and study of the Standard Model Higgs boson through its decay into a $b\bar{b}$ quark pair. It is nevertheless not well understood, and large theoretical uncertainties affect the modelling of the production of collinear b-quark pairs. Different calculation schemes are adopted by Monte Carlo generators for the simulation of the $Zb\bar{b}$ final state at the Large Hadron Collider. The analysis of the angular correlations allows to identify the simulation technique that provides the best description of the data.

The measurement is performed using a data sample corresponding to an integrated luminosity of 5.15 fb^{-1}, recorded by the CMS experiment at the LHC. The inclusive and differential cross sections are evaluated in different kinematical regions, selected by a requirement on the Z boson transverse momentum.

Collisions with a Z boson decay are selected by requiring the presence of two high momentum leptons, muons or electrons. Hadrons containing a b valence quark are identified with a technique that, being independent from jet reconstruction, extends for the first time the sensitivity to the collinear b-hadron pair production.

The measured cross sections are compared with the theoretical predictions obtained with several Monte Carlo generators, at the tree-level and at the next-to-leading order of the perturbative expansion and implementing different calculation schemes: MadGraph, Alpgen and aMC@NLO. The angular distributions are observed to be in fair agreement with the expectations within the experimental and theoretical uncertainties. The best description is provided by the Alpgen generator. Localised discrepancies are however found, particularly in the production of collinear b-hadron pairs.

The observations are consistent with the hypothesis of a mismodelling of the contribution from the $qq \to Zb\bar{b}X$ and $qg \to Zb\bar{b}X$ subprocesses, related to Feynman diagrams with a gluon splitting vertex $g \to b\bar{b}$. Moreover, a discrepancy is observed in the total cross section, the data

being approximately 15% higher than all theoretical predictions. The cross-section trend as a function of the Z boson transverse momentum is in agreement with the expectations.

This work also presents the measurement of the spatial resolution achieved by the CMS pixel barrel detector, performed with LHC collision data. The pixel hit resolution constrains the performance of track and vertex reconstruction, decisive for the identification of heavy flavour particle decays.

INTRODUCTION

The hunt for the Higgs boson has been one of the major objectives of high energy physics experiments for the last thirty years, and one of the motivations for the construction of the Large Hadron Collider (LHC) and of the Compact Muon Solenoid (CMS). In July 2012, after having analysed a proton-proton collision data sample equivalent to integrated luminosities of 5.1 fb^{-1} at $\sqrt{s} = 7$ TeV and 5.3 fb^{-1} at $\sqrt{s} = 8$ TeV, ATLAS and CMS announced the discovery of a new resonance at a mass of approximately 125 GeV [1,2]. Since then, the survey of bigger data samples and the optimisation of the analyses have lead to the confirmation of the discovery, and have allowed for a first study of the new particle properties.

In this context, the analysis of the associated production of a Higgs and a vector boson, $W^{\pm}$ or Z, with the subsequent decay of the Higgs boson into a $b\bar{b}$ quark pair, is particularly relevant. It constitutes in fact a first test of the new boson coupling to the quark sector of the Standard Model (SM) and to fermions in general.

The search and study of new physics phenomena implies an excellent understanding of the background processes. An irreducible background to the Higgs search in the $H \to b\bar{b}$ channel is the associated production of a Z boson and heavy flavour quarks. The $pp \to Zb\bar{b}X$ process is currently not well understood. Large uncertainties are in particular associated to the production of collinear b-hadron pairs, dominated by subprocesses involving Feynman diagrams with a gluon splitting vertex $g \to b\bar{b}$. The phase space region has not been probed so far, as it is inaccessible with jet-based event reconstruction techniques.

This thesis presents the measurement of the angular correlations in the $Zb\bar{b}$ final state, in other words of the differential cross section as a function of the angular separation between the b hadrons, and between the Z boson and the b hadrons, which are directly sensitive to the $Zb\bar{b}$ production mechanisms. Owing to the use of a b-hadron identification technique independent from jet reconstruction, this work for the first time extends the sensitivity to the collinear production of b-hadron pairs.

The thesis is organised as follows. Chapter 1 gives a overview of the theoretical modelling of hadronic collisions events, with a particular emphasis on the simulation of heavy quark production. The experimental setup is described in Chapter 2, while the event reconstruction

is discussed in Chapter 3. The latter chapter contains a description of a measurement of the spatial resolution achieved by the CMS pixel detector, crucial for the performance of the b-quark identification algorithms. The study of the $pp \rightarrow Zb\bar{b}X$ associated production is presented in Chapter 4.

CHAPTER

ONE

HIGGS SEARCHES AND BACKGROUND PREDICTIONS

The particle content of matter and the mechanisms of interaction between particles are described by the Standard Model (SM) [3, 4], which reproduces with very good accuracy the results of several decades of experimental measurements.

The known fundamental constituents of the universe belong to two categories: the fermionic matter particles and the bosonic mediators of their interactions. Fermions are further classified into leptons and quarks, according to the type of charges they carry. There are 6 lepton and 6 quark types grouped in pairs, constituting three generations: (e, ν_e), (μ, ν_μ) and (τ, ν_τ) for the leptons, (u, d), (c, s) and (t, b) for the quarks. Three types of interactions can occur between these particles and are accounted for by the SM: electromagnetic, between carriers of electric charge (e, μ and τ leptons and quarks), weak, between all leptons and quarks, and strong, between particles with colour charge (quarks). Interactions are mediated by the force-carrier bosons: the electromagnetic force by the massless photon γ, the weak force by the massive $W^{\pm}$ and Z, and the strong force by eight massless gluons. Electromagnetic and weak interactions are two manifestation of the same theory, and are unified in the electroweak sector.

The SM is described in the quantum field theory formalism. The theory is built by postulating a set of local gauge symmetries and by writing the most general Lagrangian that includes all the fields corresponding to the constituent particles and is invariant under these transformations. Symmetries establish the foundations of the SM, as they dictate the particle interactions, and the particle fields are representations of their groups. The SM local gauge symmetry corresponds to the group $SU(3)_C \times SU(2)_L \times U(1)_Y$. The SU(3) group refers to the strong interaction, and gives rise to the three colour charges and to the eight vector gluon fields. This part of the model is described by Quantum Chromodynamics (QCD).

In this framework, only massless bosons are accepted as force carriers, as the addition of simple mass terms for these particles would spoil the gauge invariance of the SM Lagrangian.

Nevertheless, experimental evidences show that, of the four vector bosons mediating the electroweak interactions only the photon is massless, while the Z and $W^{\pm}$ are observed to have a mass of the order of 100 GeV. This puzzle is solved by the introduction of a spontaneous symmetry breaking mechanism [5–7], which, in its minimal variant, leads to the prediction of one scalar field. The manifestation of this field is the Higgs boson (H), which couples to the fermions through the Yukawa interaction, generating their mass terms in the Lagrangian. Although constrained by theoretical argumentations as well as by direct and indirects evidences, the mass of the Higgs boson (m_{H}) is a free parameter of the SM, and has to be experimentally determined.

The Higgs boson can be directly produced at electron-positron and hadronic colliders, with a cross section that depends on the center-of-mass energy and on the boson mass. For a proton-proton machine, the Higgs production mechanisms with the largest cross section is the gluon-gluon fusion, followed by the vector boson fusion, the associated production with a vector boson (VH), and the associated production with a top quark pair, as shown in Fig. 1.1. The Higgs can decay into a wide spectrum of final states including fermions and gauge bosons. The branching ratio for each channel is determined by the H coupling to the final state particles, which depend on the boson mass. The branching fractions as a function of m_{H} are shown in Fig. 1.1.

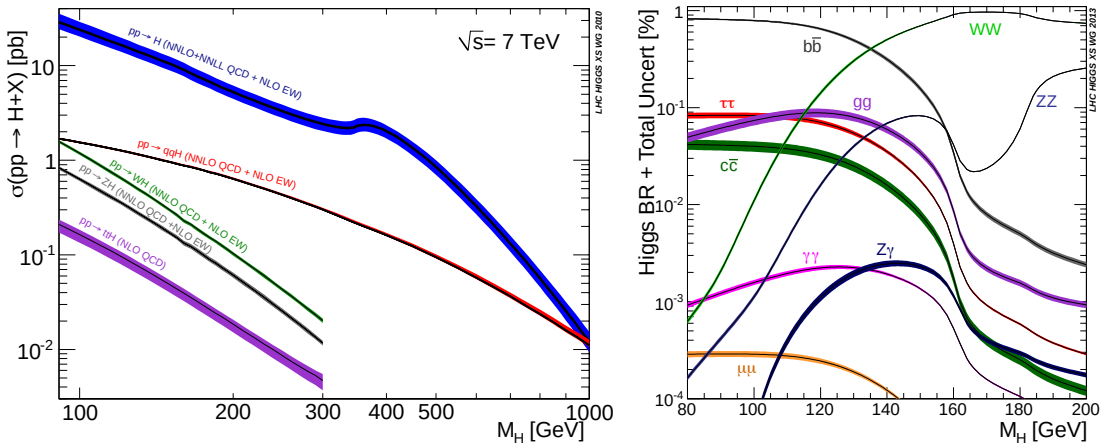

Figure 1.1: *Left*: Standard Model Higgs boson production cross section as a function of the boson mass, at a proton-proton collider at a center-of-mass energy of 7 TeV. *Right*: SM Higgs boson branching fraction as a function of the boson mass.

The hunt for the SM Higgs boson has been one of the major objectives of high energy physics experiments for the last twenty years. Searches have been performed at the LEP [8] and Tevatron [9] colliders, leading to the exclusion of the mass ranges $m_{\mathrm{H}} < 114.4$ GeV and $162 <$

$m_H < 166$ GeV, and to the observation of an excess of events with respect to the background-only expectation in the range $120 < m_H < 135$ GeV.

The search for the SM Higgs boson is also one of the priorities of the ATLAS and CMS experiments at the LHC. In July 2012, after having analysed a collision data sample equivalent to integrated luminosities of 5.1 fb^{-1} at $\sqrt{s} = 7$ TeV and 5.3 fb^{-1} at $\sqrt{s} = 8$ TeV, ATLAS and CMS announced the discovery of a new resonance at a mass of approximately 125 GeV [1,2]. Since then, the constant survey of the coming datasets and the optimisation of the analyses have lead to the confirmation of the discovery, and have allowed for a first study of the new particle properties aimed at determining its nature. Several search channels are explored, in the attempt to cover all Higgs production modes and decay final states. At the time of this work, significant excesses of events with respect to the background-only hypothesis are found in the most sensitive decay channels $H \to \gamma\gamma$, $H \to ZZ$ and $H \to WW$, while hints of a signal appear in the channels $H \to b\bar{b}$ and $H \to \tau\tau$. The signal strength — defined as the ratio σ/σ_{SM} between the observed cross section and that predicted by the SM — for all the search channels combined is compatible with one. The ratio between the observed cross section and that predicted for a SM Higgs boson for the various search channels is shown in Fig. 1.2.

The measurement of the production cross section in each decay channel is crucial for determining the particle properties: the observation of its decay into a photon pair indicates that it is a boson, while the branching ratios are proportional to the couplings to the different final state particle types.

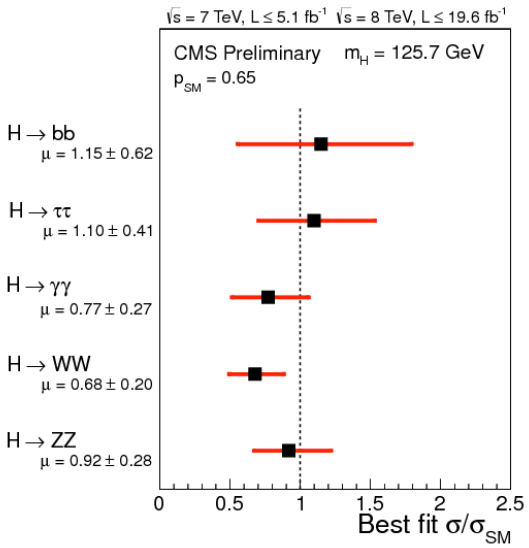

Figure 1.2: Cross sections of the newly discovered boson measured in different final states, divided by the expected value for a SM Higgs boson.

The study of the $H \to b\bar{b}$ decay [10] is particularly relevant, as it constitutes the first test of the boson coupling to the quark sector, and, together with $H \to \tau\tau$, to fermions in general. Despite the high branching ratio in the low m_H region, the observation of the $H \to b\bar{b}$ decay is not straightforward. The search for the higher cross-section process $gg \to H \to b\bar{b}$ is unfortunately prohibitive, due to the overwhelming backgrounds from b-quark pair production. Instead, it is convenient to exploit the VH production mechanism, $q\bar{q} \to VH \to Vb\bar{b}$.

One of the main sources of background to this final state is the QCD production of a vector boson, W or Z, is association with a b-quark pair ($pp \to W^{\pm}b\bar{b}X$ and $pp \to Zb\bar{b}X$). The estimation of the background contamination and the extraction of the signal yield are performed with a multi-variate technique, relying on event samples simulated by Monte Carlo (MC) generators. The predictions are validated by comparing them with the data in several control regions, highly dominated by a single process. For the $Zb\bar{b}$ final state, this comparison reveals a significant discrepancy, with the data being approximately a factor of two higher than the expectation [10]. The disagreement is observed to be mostly related to the production of collinear b-quark pairs. To reduce the impact of a wrong background estimation in the Higgs search, the predicted $Zb\bar{b}$ component is rescaled to reproduce the data. This reweighting constitutes the source of one of the largest systematic uncertainties, which could be suppressed by improving the goodness of the modelling of the $Zb\bar{b}$ production mechanisms.

The collinear b-quark pair production in association with a Z boson is in general not well understood. The large uncertainties in this context are partly related to the difficulty to probe this region of the phase space with the common experimental techniques. Although the total $Zb\bar{b}$ production cross section has been measured by the ATLAS and CMS experiments at LHC [11, 12], all analyses exploit jet reconstruction, which is not adequate to detect parton pairs at small angular separations. The angular correlation analysis discussed in this thesis for the first time extends the measured phase space, shedding light on all the $Zb\bar{b}$ production mechanisms, therefore helping improving the reliability of the theoretical predictions.

This chapter presents the ideas behind the development of theoretical predictions for physics processes at a hadronic collider, with a particular emphasis on the heavy flavour production. Section 1.1 gives a general overview of the modelling of a hadronic collision event, based on the factorisation of hard and soft components, while a more detailed description of the techniques to simulate the hard interaction subprocess is given in Section 1.3. The various tools currently available are also presented.

1.1 High energy processes, hadron collisions

A hadronic collision is a phenomenon of great complexity [13]. The evolution of a collision event can be summarised as follows:

1. The two beams collide in the intersections points, and pairs of beam particles interact.

Each hadron contains quarks of different flavours and gluons, which carry fractions of its momentum. The hadron composition is terms of flavour and energy sharing is modelled by parton distribution functions (PDF);

2. The partons embedded in the colliding hadrons emit radiation, initiating a sequence of branching processes $q \to qg$, $g \to q\bar{q}$ and $g \to gg$. Due to the large value of the strong coupling constant α_S, these splittings have a high probability to occur, resulting in the formation of initial-state parton cascades;

3. Two partons in the cascades enter the hard interaction, at a momentum transfer scale Q^2. The products of the hard scattering are the final-state elementary particles, partons, leptons and bosons, that characterise the event topology. Short-lived resonances, such as Z, $W^{\pm}$ and Higgs bosons, immediately decay into partons, leptons or photons. Although not observable, the hard scattering subprocess determines the main properties of the collision event;

4. The outgoing partons start branching, and initiate the development of final-state cascades;

5. At each branching in the initial and final-state showers, the momentum scale decreases, down to the cutoff scale $\Lambda_{QCD} \sim 1$ GeV, at which the perturbative theory loses its validity;

6. Below Λ_{QCD}, the strong interaction confines the partons into colourless hadrons. The confinement process is known as fragmentation, and is followed by the decay of the unstable particles. Through fragmentation and decay, which together form the hadronisation, the parton cascades evolve into jets of stable and meta-stable particles that constitute the observable final state of a collision event.

A graphical representation of this evolution is given in Fig. 1.3.

The modelling of a hadronic collision as a whole is a task of prohibitive proportions. The solution is to factorize the evolution into subprocesses, each of them relatively easier to handle with the appropriate technique. This approach is adopted in the Monte Carlo generators, the main tool to describe and reproduce the phenomenology at a hadronic collider.

The *factorisation theorem* justifies the independent treatment of the hard scattering and of the soft non-perturbative processes. For a hadronic (proton-proton) collision $p_A p_B \to X$, where X is a generic final state, it is expressed by the formula:

$$\sigma_{AB} = \int dx_a dx_b f_a(x_a, Q^2) f_b(x_b, Q^2) \cdot \hat{\sigma}_{ab \to X} \tag{1.1}$$

where σ_{AB} is the total cross-section, x_a and x_b are the fractions of the proton momentum carried by the two partons a and b involved in the interaction, and $\hat{\sigma}_{ab \to X}$ is the hard partonic scattering cross section. The latter can be calculated with the perturbative expansion:

$$\sigma_{AB} = \int dx_a dx_b f_a(x_a, \mu_F^2) f_b(x_b, \mu_F^2) \cdot [\hat{\sigma}_0 + \alpha_s(\mu_R^2)\hat{\sigma}_1 + \alpha_s^2(\mu_R^2)\hat{\sigma}_2 + ...]_{ab \to X} \tag{1.2}$$

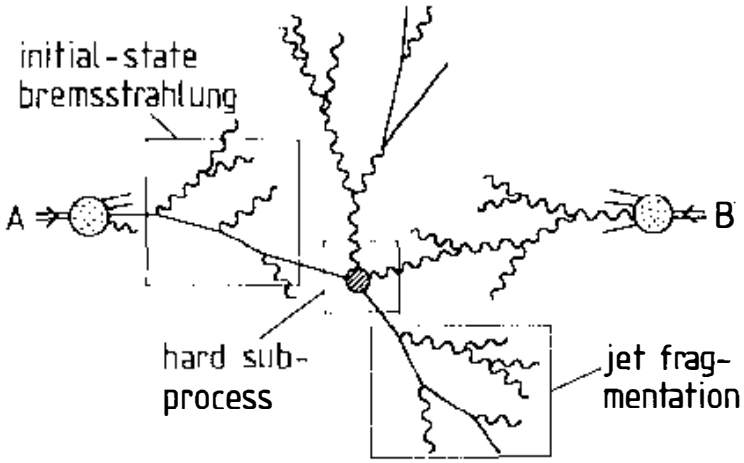

Figure 1.3: Hard scattering and initial and final state radiation, in a hadronic collisions producing two jets.

The momentum transfer Q^2 is replaced by the *factorisation scale* parameter μ_F, which identifies the separation between the hard scattering and the soft process. In the perturbative expansion, the strong coupling α_S is evaluated at the *renormalisation scale* μ_R. The two scale parameters are unphysical. At all perturbative orders, the μ_F and μ_R dependence of the parton density functions and of α_S is exactly compensated by the $\hat{\sigma}_i$ coefficients, resulting in the invariance of the σ_{AB} cross section under changes of their values. At a fixed order, instead, the dependences do not cancel out, and a specific choice of the scale parameter values is necessary for a cross-section estimation. The sensitivity of the σ_{AB} prediction to variations of the scale parameters has to be accounted for as theoretical uncertainty. The $f_a(x_a, \mu_F^2)$ and $f_b(x_b, \mu_F^2)$ terms are the parton distribution functions, described in Section 1.2.

Qualitatively, the factorisation theorem states that the hard scattering and the soft components of a hadronic collision can be disentangled, and independently modelled. The hard-scattering component for a specific process of interest can be solved in the context of the well known perturbation theory. Different methods are available, as described in Section 1.3. Conversely, part of the soft process, namely hadronisation and quark confinement, occurs at the momentum scales of non-perturbative QCD. Phenomenological models have also to be used in this context, based on experimental data. Because of the process-independence of the soft-process phenomenology, these models have a general validity.

1.2 Parton distribution functions

The PDF functions in Eq. 1.1 and 1.2 relate the dynamics of the partons participating in the hard scattering to that of the colliding hadrons, by modelling the probability that a parton

carries a fraction x of the total hadron momentum. The PDF show a dependence on the momentum transfer Q^2, which is generated by higher-order corrections from real and virtual gluon emission within the colliding protons. In good approximation, the Q^2 evolution is determined by the Altarelli-Parisi, or DGLAP, equations, developed in the perturbative theory. The x dependence is extracted from a global fit of data, including few thousands measurement points from deep inelastic scattering (DIS), Drell-Yan and jet production. Results are available at the tree-level, at the next-to-leading (NLO) and, only partially, at the next-to-next-to-leading (NNLO) order. The PDF extrapolation is affected by an uncertainty, constrained by the accuracy of the experimental data, by the uncertainty on the coupling α_S, and by the accuracy of the analysis, and has to be taken into account as source of uncertainty in all theoretical predictions.

1.3 Modelling of the hard process

This section presents the main approaches adopted in the Monte Carlo generators to reproduce the hard subprocess in a hadronic collision: the matrix element calculation, and the *parton shower*. A third method, merging the two techniques, which is at present extensively used by the main MC programs, is also discussed.

1.3.1 Matrix element calculation

The traditional and most natural way to study a hard scattering process is to compute the squared amplitudes of the corresponding Feynman diagrams, order by order, and to integrate them over the appropriate phase space. This approach allows to take into account interference effects, and to correctly reproduce kinematics and helicity structures.

Due to the technical complexity of the matrix element calculations, predictions obtained at the leading order of the perturbative expansion are most common. In most cases, the phase-space integration is not analytically solvable, and is carried out with numerical techniques. Tree-level calculations are extensively used at the experiments, since they provide a reasonably good description of the event structure for many processes.

The tree-level approximation is however not optimal. As mentioned in Section 1.1, lowest-order calculations depend on the choice of the unphysical factorisation and renormalisation scales, resulting in large uncertainties on the theoretical predictions. This problems can be solved by extending the perturbative expansion to the higher order in the strong coupling α_S, including additional radiation of quarks and gluons, and loop corrections.

A bare matrix-element calculation reliably describes the fundamental parton interaction, characterised by hard momentum scales, but does not correctly handle the emission of soft and collinear radiation. In addition, a complete matrix element calculation becomes rapidly im-

possible as the final-state parton multiplicity increases. This method is therefore inadequate to model the sequence of parton branchings leading to the formation of the cascades observed at the experiments. Alternative techniques are available for this purpose, as discussed in the next section.

1.3.2 Parton shower

The showering model allows to reproduce the formation of parton cascades from the hard-process scale to the cutoff scale Λ_{QCD}, where non-perturbative QCD takes over.

A shower consists of a sequence of parton splittings a $\rightarrow$ bc, of the type q $\rightarrow$ qg, g $\rightarrow$q$\bar{\text{q}}$ or g $\rightarrow$ gg, for each of the coloured partons in the initial and final state. At each branching, part of the energy of the primary parton is carried away, and the momentum of the partons in the shower becomes softer. This progression is modeled using an *evolution variable* (t), which takes values between the hard scale t_0 and the cutoff t_{cut}. The cascade begins with the few final-state partons at the hard subprocess scale t_0. Each of them has a certain probability of emitting a parton carrying away a fraction z of its energy — depending on z and on the type of partons involved — according to the Altarelli-Parisi splitting functions $P_{ba}(z)$ [14]. After the branching, both partons can split again at a softer scale.

To practically handle the branching in the context of a shower generator, it is convenient to use the *Sudakov form factors* formalism:

$$\Delta(t_1, t_0) \equiv \exp(-\int_{t_0}^{t_1} \frac{dt}{t} \int \frac{dz}{z} \frac{\alpha_S}{2\pi} P_{ba}(z)). \tag{1.3}$$

The Sudakov factors model the probability for a parton to evolve from the scale t_0 to the softer scale t_1 without radiating a parton with energy fraction z. The integration is extended to the phase space region where the emission is resolvable. Outside that range, the radiated object is either too soft — softer than the non-perturbative cutoff — or too collinear to the primary parton to generate an observable jet.

A Monte Carlo generator therefore proceeds as follows: at the hard scattering scale t_0, the program extracts a value of the evolution variable t_i by solving the equation $\Delta(t_i, t_0) = R_i$, where R_i is a random number between 0 and 1. If $t_i < t_{cut}$ the emission is unresolvable, and the shower of that primary parton is terminated. Conversely, if $t_i > t_{cut}$, the splitting occurs. The procedure is repeated iteratively until the scale reaches the cutoff. At this stage, the hadronisation occurs.

The discussion above refers to the modelling of a final-state shower. A similar technique is adopted for the description of initial-state parton cascades. An initial-state shower is initiated by the partons embedded in the incoming hadrons, and is terminated at the scale at which the hard scattering occurs. The most suitable approach to reproduce this process is the so-called *backward evolution*. The first step is to choose the two partons involved in the hard scattering,

and their momentum fraction x. The configuration is subsequently evolved backwards through sequential branchings at increasing energy scale, using the same evolution variable as for the final state shower, and modified Sudakov form factors that include the PDF.

Common shower programs in use at LHC experiments are PYTHIA [15] and HERWIG [16]. The main differences between the two generators are the evolution variable t, by default the parton virtuality for PYTHIA, and a combination of the parton energy and the emission angle for HERWIG, and the choice of the hadronisation model, the string in PYTHIA and the cluster in HERWIG.

1.3.3 Combining matrix element and parton shower: the jet matching

As discussed above, the matrix element and parton shower techniques are suitable for modelling different configurations, the matrix element calculation being the most reliable for the hard scattering subprocess, and the parton shower being more appropriate for soft and collinear radiation emissions and jet formation. Under these premises, it is clear that a combination (matching) of the two approaches would be the best solution for the extraction of theoretical predictions to compare to hadron collider data. The discussion below is valid explicitly for the matching in tree-level Monte Carlo generators. The details of the procedure adopted for the matching of the parton shower to next-to-leading-order calculations are given in [17].

The main obstacle to the practical implementation of the matching is the definition of the separation between the hard component of the process, to be solved with the matrix element calculation, and the soft and collinear emissions that instead compete to the parton shower. There is in fact an intrinsic ambiguity: an event with a certain number N of jets can be produced either from the showering of the appropriate N-parton final state, or from an $(N-1)$-parton final state, with the radiation of an additional hard parton during the shower evolution. The two approaches are equivalent, and lead to the same result. Factorisation theorems are unfortunately not rigorously applicable to complex final state topologies, which are characterised by several hard scales.

A specific factorisation prescription — the *matching scheme* — is introduced, identifying on an event-by-event basis the approach that provides the best description of a given configuration. The matching scheme has to avoid double counting in the phase-space regions of overlap between the matrix element calculation and the parton shower, as well as dead zones. The most natural solution is to apply a cutoff, known as *matching scale*: branchings occurring at a scale harder than the cutoff are handled by the matrix element calculation, while softer radiation emissions are left to the shower program. As the matching scale is unphysical, the resulting theoretical predictions should not depend, or at least show a small dependence on the choice of the cutoff.

Several matching schemes have been developed, such as the MLM [18] and the CKKW [19]. Only The MLM approach is described in detail here, as it is adopted by all the generators used in the analysis forming the main topic of this thesis.

The MLM jet matching scheme requires the definition of a measure, i.e. a variable describing the distance between two objects in terms of momentum and angle. As the name suggests, the k_T-MLM method uses the k_T measure defined in [20]. An alternative choice, not discussed here, is given by the transverse momentum and the three-dimensional angular separation. Events are generated in a sequence of several steps:

1. The matrix element calculation is performed. The squared amplitudes are computed automatically, taking into account all the diagrams with parton multiplicities below a maximal value N^{max}_{parton}. A cutoff Q_{ME} is applied at this stage, so that only partons with $k_T > Q_{ME}$ are generated. The cutoff is usually chosen to be significantly lower than the factorisation scale μ_F, such that a substantial fraction of the emissions are classified as hard/well separated, and produced by the matrix element calculations. In order to ensure a smooth transition between the matrix element and the parton shower, the strong coupling at each splitting vertex is evaluated at the k_T at which the branching occurs, $\alpha_S(k_T)$. This is known as α_S *reweighting*;

2. The parton shower takes as input the event generated by the matrix element, and creates additional radiation down to the perturbative cutoff at $\Lambda_{QCD} \sim 1$ GeV;

3. The event is clustered into jets using the k_T measure as criterion, with a constraint $k_T < Q_{cut}$. The Q_{cut} parameter defines the jet resolution scale below which the parton shower takes over. The Q_{cut} and Q_{ME} values in general do not coincide, as they are applied at different stages of the emission evolution, and have to be chosen such that to exclude double counting of event topologies as well as dead zones;

4. The double counting is removed by applying an event-by-event veto. A single event is retained if:

   - $N_{jet} = N_{parton}$, where N_{jet} is the number of resolvable jets, i.e. jets with $k_T > Q_{cut}$, or in other words if each resolvable jet produced by the parton shower is matched to a matrix element parton;
   - if $N_{jet} > N_{parton}$, only in the highest jet multiplicity sample, and if the k_T of the parton shower jets is lower than the k_T of the softest jet matched to a matrix element parton. This condition allows for the production of events with all jet multiplicities.

In the MLM scheme, letting the shower generate parton emissions at any scale and then vetoing the undesired events may consistently suppress the efficiency of the generation process. An option is given by the CKKM: it avoids the veto approach, and adopts a more sophisticated

event reweighting technique based on Sudakov form factors [19]. The main disadvantage is the complexity of the algorithm, making it technically difficult to implement in a generator.

1.3.4 Generation of processes involving b quarks

The previous sections introduced the basics of hadronic collision event generation. The present section focusses on the modelling of the heavy quark and jet production. Due to their heavier mass, $m_b \sim 4.7$ GeV, the treatment of event topologies containing b quarks in the initial state is not trivial. Two approaches can be adopted, known as four (4F) and five flavour (5F) according to the number of parton flavours included in the initial state [21].

The 5F scheme is based on the assumption that the b quark is massless, assimilating it to the lighter parton types u, d, c, s-quark and gluon (g). This approximation is reasonable as long as the hard scale of the process is significantly higher than m_b. The whole generation sequence treats b quarks as any other type of parton: the logarithmic divergencies in the initial state are resummed into a b-quark PDF, added to the u, d, c, s, g functions; the jet matching procedure described above is also applied to b quarks, with the same matching scales as for the lighter partons. This simplifies the technical implementation of the calculations.

The 4F scheme on the contrary considers the b quark massive, decoupling it from the lighter parton types. The heavy mass provides for a natural cutoff in the parton splittings involving b quarks and protects from infrared and soft/collinear divergencies, which do not need to be resummed in a b-quark PDF. As a result, b-quarks are not part of the initial state, and their production has to be accounted for explicitly in the matrix element calculation via the gluon splitting $g \rightarrow b\bar{b}$. An example of that is given in Fig. 1.4, which shows the different interpretation of the same leading-order $Zb\bar{b}$ production subprocess in the four and in the five-flavour schemes. The massive b-quark assumption also implies the exclusion of b-quarks from the jet matching. For this purpose, in the k_T-MLM scheme the b-quarks and all the partons radiated from them are removed from the list of showered partons constituting the input to the k_T clustering. This choice is justified under the hypothesis that no resolvable jet originates from the b-quark or from its products.

The b mass constitutes a protection against divergencies only at relatively soft scales. As soon as the hard process scale is $Q^2 >> m_b$, large initial state logarithms may arise. These are not resummed in the b-quark PDF, and can therefore lead to a poor behaviour of the perturbative expansion. However, such effects are in general smaller than other approximations [21].

To all perturbative orders, the four and five-flavour schemes can give identical predictions. At finite order, the different ordering of the perturbative expansion may result in considerable differences in the properties of the generated events.

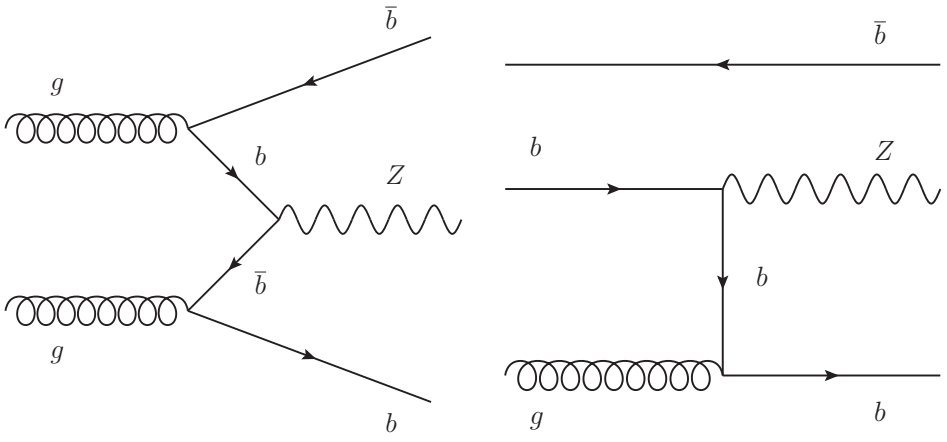

Figure 1.4: Feynman diagrams corresponding to different approaches in describing a leading-order $Zb\bar{b}$ production subprocess: in the four-flavour scheme (*left*) the b-quarks are explicitly created in the context of the matrix element calculation, while in the five-flavour approach (*right*) their creation is embedded in a b-quark PDF.

1.3.5 Generators for the $Zb\bar{b}$ associated production

The methods described in the previous sections are implemented in various Monte Carlo generators used at hadron collider experiments. This section gives an overview of those considered in the analysis presented in this thesis.

For the measurement of the $Zb\bar{b}$ production cross section, the reference MC generator is MadGraph [22, 23]. It implements an automated calculation of the scattering amplitudes from all Feynman diagrams contributing at the tree-level to the process of interest, and performs the phase-space integration with Monte Carlo techniques. MadGraph allows for high final state parton — or jets — multiplicities. Both four and five-flavour schemes can adopted, with the b-quark mass set to 4.7 GeV in the 4F. For the study of the $Zb\bar{b}$ associated production the five-flavour approach has a technical disadvantage: it generates events with a Z boson is association with jets of any flavour, according to the corresponding cross section, which heavily limits the size of a $Zb\bar{b}$ sample. The adoption of the 4F scheme instead allows for the production of events with the Z in association with b quarks. In MadGraph the matrix-element partons are matched to a parton shower, handled by the Pythia generator. The k_T-MLM matching scheme is adopted, with the matching scale set to 30 GeV for the 4F and to 20 GeV for the 5F. The CTEQ5L1 PDF set is used for the 4F, while the 5F implements the CTEQ6L1 set.

Generators not relying on the explicit calculation of Feynman diagrams have also been developed. The Alpgen leading-order matrix-element generator implements the alternative method described in [24]. Similarly to MadGraph, it allows for high final-state parton multiplicities.

The $Zb\bar{b}$ ALPGEN available sample is produced in the four-flavour approach, adopting the k_T-MLM matching scheme.

The generators mentioned above perform the matrix element calculation at the leading order of the perturbative expansion. In most cases, the tree-level approximation is good enough to obtain a reliable description of the event structure. However, for the reasons mentioned in Section 1.3.1, it is in general appropriate to compare the data to a prediction that includes next-to-leading-order corrections. A recently developed NLO tool is the aMC@NLO generator [25], which implements a fully automated calculation of the amplitudes from the Feynman diagrams at NLO precision, under the MADGRAPH formalism. The b quark is assumed to be massive. The matrix element calculation is matched to the HERWIG or PYTHIA parton showers as described in [26].

The choices of factorisation and renormalisation scales adopted in the various generators are summarised in Table 1.1. The four and five-flavour MADGRAPH generators implement the CTEQ5L and CTEQ6L respectively, while ALPGEN uses the MSTW2008LO set. The aMC@NLO generator adopts the next-to-leading-order MSTW2008.

	μ_F^2	μ_R^2
MG 5F	$m_Z^2 + p_T^2(\text{jets})$	k_T^2 at each vertex splitting
MG 4F	$m_Z^2 + k_T^2(\text{b},\text{b})$	k_T^2 at each vertex splitting
ALPGEN	$m_Z^2 + \sum_{\text{jets}}(m_{\text{jets}}^2 + p_{T,\text{jets}}^2)$	k_T^2 at each vertex splitting
aMC@NLO	$m_{\ell\ell'}^2 + p_T^2(\ell\ell') + \frac{m_{\text{b}}^2+p_T^2(\text{b})}{2} + \frac{m_{\text{b}}'^2+p_T^2(\text{b}')}{2}$	$= \mu_F^2$

Table 1.1: Scale choices in the different theoretical predictions for the factorisation (μ_F^2) and renormalisation (μ_R^2) scales.

1.4 Below Λ_{QCD}: fragmentation and decay

The previous sections explored the techniques to treat the hard scattering subprocess of a hadronic collision, valid as long as the momentum scale is above the Λ_{QCD} cutoff. Below, the process cannot be described in the perturbative framework. Several phenomenological models have been developed to simulate the transition between the unobservable partonic final state, and the jets of particles experimentally detectable. The three main available approaches are the independent fragmentation, substantially outdated, the *cluster* [27], and the *string* [28], implemented in different generators.

The hadronisation phase also includes the decay of all unstable hadrons and leptons produced by the fragmentation. This requires the implementation in the generator of the appropriate mass, decay width and branching ratio for each type of particle, collected in decay tables.

Polarization and, for bottom and charmed hadrons decays, the proper elements of the Cabibbo-Kobayashi-Maskawa matrix have to be taken into account.

1.5 Underlying event and multiple parton interactions

When two partons participate in the hard interaction, they leave behind the other constituents of the colliding hadrons. These partons — the beam remnants — take part in the hadronisation and contribute to the event final state. This activity is known as *underlying event*. There is a non-negligible probability of multiple hard interactions (MPI) between partons within the same beam hadrons. A study of the $Zb\bar{b}$ production mechanisms should therefore take into account the possibility that the Z boson and the b-quark pair are products of two independent parton scatterings.

An approximate estimate of this contribution to the $pp \to Zb\bar{b}X$ cross section measured in this work can be obtained with the formula:

$$\sigma_{MPI}(Zb\bar{b}) = \frac{\sigma(Z) \times \sigma(b\bar{b})}{\sigma_{eff}} \tag{1.4}$$

where σ_{eff} accounts for the proton structure [29], while $\sigma(Z)$ and $\sigma(b\bar{b})$ are the cross sections for the inclusive Z production $pp \to ZX$ and for the $pp \to b\bar{b}X$ process. From experimental results and Monte Carlo simulation, these are evaluated to be $\sigma_{eff} \sim 15$ mb, $\sigma(Z) \sim 0.4$ pb and $\sigma(b\bar{b}) \sim 2.2$ µb in the phase space considered for the analysis presented in this thesis, defined by the requirements:

- Lepton $p_T^\ell > 20$ GeV and $|\eta^\ell| < 2.4$, for the leptons from the Z boson decay;
- Dilepton invariant mass $81 < M_{\ell\ell} < 101$ GeV;
- b-hadron $p_T^B > 15$ GeV and $|\eta^B| < 2.0$.

The resulting cross section for the MPI $Zb\bar{b}$ production is approximately 2 fb, about 1% of the single-scattering $pp \to Zb\bar{b}X$.

1.6 Summary

This chapter explained the motivations for a measurement of the cross section and angular correlations of the $Zb\bar{b}$ associated production at the LHC, by exploring the theoretical framework for the modelling of hadronic collision events. A particular attention was addressed to the various techniques developed for the simulation of QCD processes involving heavy-flavour particles: the five-flavour scheme, based on the massless b-quark approximation, and

the four-flavour scheme, which considers the b quark massive. At finite order in the perturbative expansion the predictions obtained in the two approaches may significantly differ. The work presented in this thesis is aimed at identifying the approximation that best describes the data.

CHAPTER

TWO

THE LARGE HADRON COLLIDER AND THE CMS EXPERIMENT

The measurement presented in this thesis exploits data from proton-proton collisions produced by the Large Hadron Collider and recorded by the Compact Muon Solenoid. This chapter gives an overview of the CERN accelerator complex and of the CMS experiment. The discussion is focalised on the subdetector systems that are most relevant for this work.

2.1 The Large Hadron Collider

The Large Hadron Collider (LHC) at CERN is the largest and one of the most complex scientific instruments ever built [30]. Its 26.7 km ring is designed to accelerate proton beams up to the unprecedented energy of 7 TeV, for a total center-of-mass energy $\sqrt{s} = 14$ TeV.

The machine started running in September 2008, followed by an accident that suspended the data taking for more than one year. For safety reasons, at the restart in November 2009 the machine was cautiously operated at much lower energy and intensity, and a long development is still needed before the design performance can be achieved.

The LHC is part of a complex of linear and ring particle accelerators that gradually increase the proton energy, some of which have been in operation since the late 1950's. The first stage is the LINAC 2 linear machine, which boosts the protons to an energy of 50 MeV and transfers them to a sequence of ring accelerators: the Proton Synchrotron Booster, the Proton Synchrotron (PS), and the Super Proton Synchrotron (SPS). Protons are finally injected from two directions into the LHC ring, where they receive a boost of 0.5 MeV per turn. As soon as the maximum energy is reached, the two proton beams are brought to collision in the four crossing spots where the CMS, ATLAS [31], LHCb [32] and ALICE [33] experiments are installed. An overview of the CERN accelerator complex is shown in Fig. 2.1.

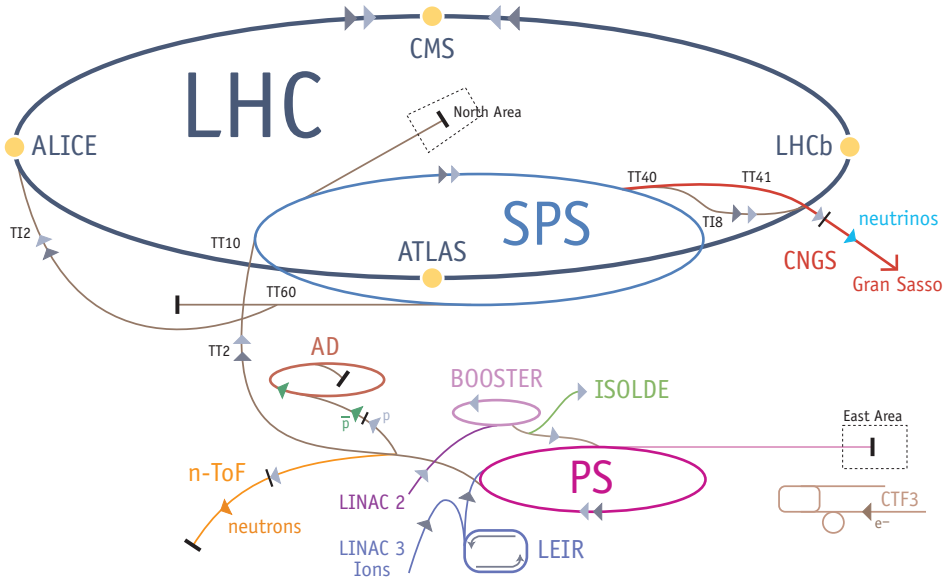

Figure 2.1: Schematic overview of the CERN accelerator complex.

The LHC accelerator is housed in the underground tunnel previously occupied by the LEP machine [34]. The two beams are accelerated by 16 superconducting radio frequency cavities (RF) and bent by 1 200 superconducting dipole magnets, cooled down to 1.9 K with supra-fluid helium and carrying the high electrical currents — up to 11 850 A — needed to provide the required magnetic field. In addition, 10 000 superconducting magnets are used for steering and focusing. An advanced cryogenic system takes care of maintaining the components in the superconducting regime.

Parameter	Design value	Best value achieved
Beam energy	7 TeV	4 TeV
Number of protons per bunch	1.15×10^{11}	1.5×10^{11}
Number of bunches	2808	1368
Crossing angle	300 µm	290 µm
Beam size	17 µm	20 µm
Emittance	3.75 µm	2.4 µm
Peak luminosity	10^{34} cm^{-2}s^{-1}	7.5×10^{33} cm^{-2}s^{-1}

Table 2.1: Relevant LHC machine parameters. The design values are compared to the ones reached at the end of the 2013 operations.

The use of radio frequency cavities forces the beams into a *bunch* structure. The LHC machine is designed to contain up to 2808 bunches of 10^{11} protons each, leading to a spacing of 25 ns and a bunch crossing rate at the experiments of about 40 MHz. A set of relevant machine

parameters is listed in Table 2.1. The nominal values, shown in the left column, have not been reached yet. Nevertheless, three years of constant efforts in machine development from the beginning of the operations at low energy and low intensity allowed to achieve most of the planned physics goals.

The number of protons per bunch, the bunch crossing frequency and the beam focalisation altogether determine the *instantaneous luminosity* — which measures the number of collisions produced per unit of surface and time — as well as the number of proton-proton hard scattering interactions at each bunch crossing (*pile up*). The instantaneous luminosity can be obtained from

$$\mathcal{L} = \frac{N_p^2 n_b f_{rev} \gamma_r}{4\pi\epsilon_n \beta^*} F \tag{2.1}$$

where N_p and n_b are the number of protons per bunch and the total number of bunches respectively, f_{rev} is the rotation frequency, ϵ_n and β^* describe the beam focalisation at the interaction points, γ_r is a relativistic factor, and F a crossing angle factor. The peak instantaneous luminosity that corresponds to the nominal machine parameters is 10^{34} cm^{-2}s^{-1}, with a pile up of 25. Despite a smaller number of bunches, the highest value registered in the 2009-2012 run is 7.5×10^{33} cm^{-2}s^{-1}, only 30% lower than the design one. Such high intensity was achieved by exceeding the design number of protons per bunch, and by improving the focussing of the beams before their collision, with a price to pay in terms of a higher pile up which makes the event selection and reconstruction more difficult. The number of interactions per bunch crossing registered in each proton fill of the LHC machine is shown in Fig. 2.2.

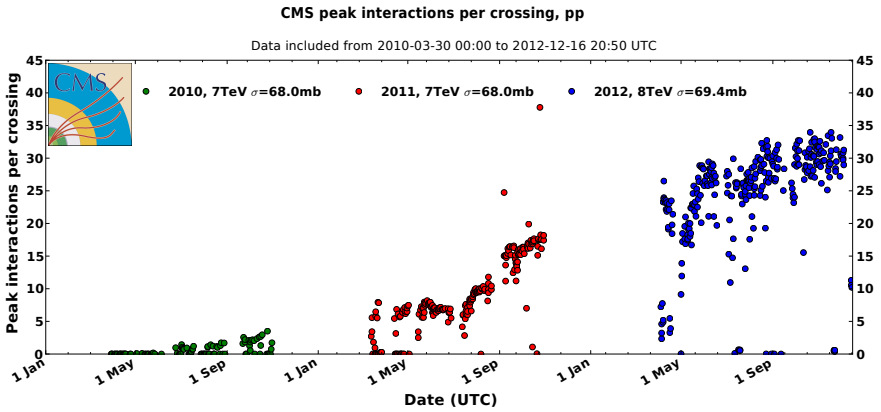

Figure 2.2: Average number of primary interactions per bunch crossing in every proton fill, registered by the CMS experiment between 2010 and 2012.

Thanks to this effort, the LHC has delivered to the experiments an enormous amount of data. Integrated luminosities of 5.15 fb^{-1} and 23.5 fb^{-1} were collected in 2011 and 2012, at $\sqrt{s}$ = 7 TeV and 8 TeV respectively, where the integrated luminosity is defined as the integral of

the instantaneous luminosity over the data taking time. The integrated luminosity collected between 2010 and 2012 is shown in Fig. 2.3.

This thesis describes a measurement performed using the 2011 collision data sample collected by the CMS experiment.

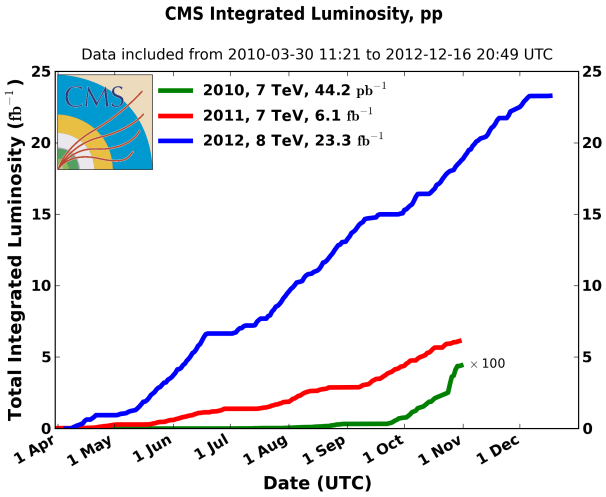

Figure 2.3: Integrated luminosity of the proton-proton collision data samples collected by the CMS experiment between 2010 and 2012.

2.2 The general concept of the CMS detector

The Compact Muon Solenoid experiment [35,36] is a general-purpose detector, designed to be sensitive to a wide spectrum of unknown physics phenomena [37]. Its ambitious research plan is highly demanding in terms of detector and event reconstruction performance. It requires high efficiency and excellent resolution in the reconstruction of the charged particle trajectories in the proximity of the proton interaction region, a precise estimation of the energy of electrons, photons, and neutral hadrons, and good muon identification up to very high energy scales. Reaching such standards is particularly challenging in the environment of high luminosity hadron collisions, producing from a hundred to few thousands final-state particles at each bunch crossing.

The key feature of the CMS experiment is its intense magnetic field, as strong as 3.8 T, which dictates the design of the entire detector. It bends the trajectories of charged particles, allowing for a precise measurement of their momentum up to energies of the order of few TeV. The field is generated by a superconducting solenoid of 13 m of length and 6 m of diameter, that saturates a 1.5 m iron yoke [38,39]. Three different types of gaseous chambers are integrated in the yoke. The cavity of the magnet coil accommodates a system of silicon detectors — the *tracker*

— located in the proximity of the interaction region and used for the precise reconstruction of the charged particle trajectories. The tracker is surrounded by an electromagnetic calorimeter, where photons and electrons are detected through the production of electromagnetic showers, and by a hadronic calorimeter, which absorbs charged and neutral hadrons to measure their energy. Only muons traverse the entire thickness and penetrate the outer gaseous tracking detectors, which are hence useful for particle identification. The structure of one sector of the CMS detector is shown schematically in Fig. 2.4. More details about each subsystem are given in the following sections.

The CMS conventional coordinate system and experimental observables

The conventional global coordinate system of CMS has the origin located in the nominal beam intersection point at the center of the detector, the x axis pointing radially towards the center of the LHC ring, the y axis pointing vertically upwards, and the z axis pointing tangentially along the beams towards the Jura mountains. In the transverse plane, the orientation with respect to the x axis is described by the azimuthal angle ϕ, while the polar angle θ measures the inclination with respect to the positive z direction. θ is commonly replaced by the *pseudorapidity*, defined as $\eta = -\ln\tan\theta/2$, which has the advantage of being Lorentz invariant. In addition to the global coordinate system, a local reference frame is used within each subdetector. A widely used observable is the projection of the momentum of a particle onto the transverse plane, indicated as p_T, which at hadron colliders is of greater physical interest than the total momentum p. Three-dimensional angular distances between objects are usually described through the quantity $\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2}$, where $\Delta\phi$ and $\Delta\eta$ are the azimuthal and pseudorapidity separations.

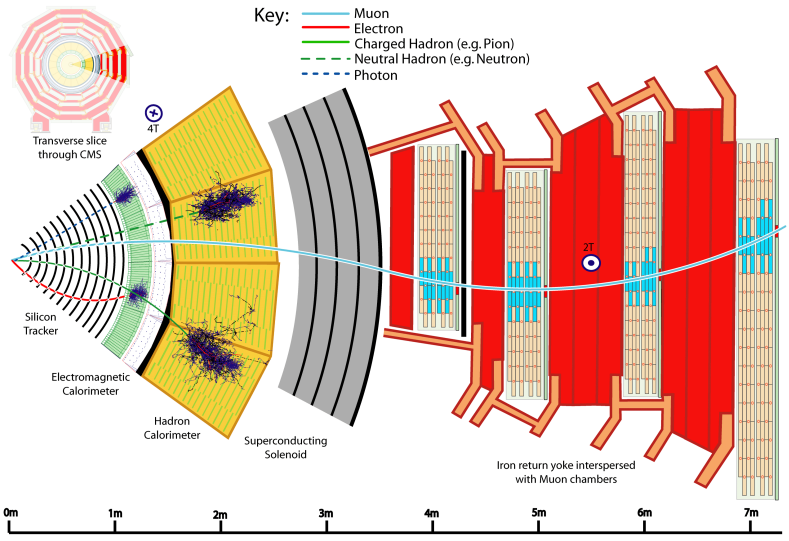

Figure 2.4: Schematic view of a transverse sector of the CMS detector.

2.3 The inner tracking system

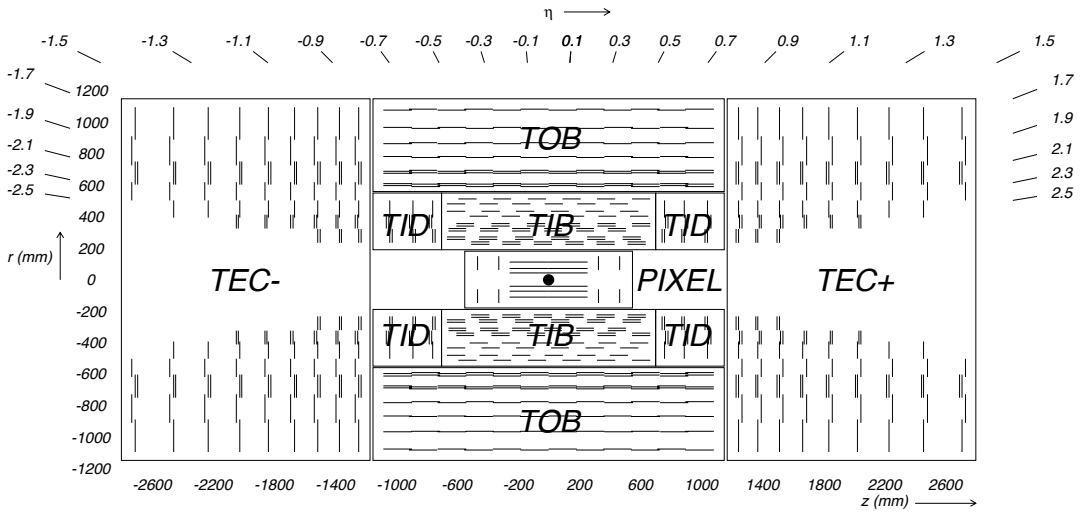

Figure 2.5: Longitudinal cross section of the CMS tracker, with the corresponding pseudorapidity coverage.

The CMS tracker [40] has the crucial role of providing the series of spatial measurement points (*hits*) that is the input to the reconstruction of charged particle trajectories (*tracks*) in the proximity of the primary interaction. Tracks contain all the information about the particle position, propagation direction and momentum, and are used to identify the primary interaction vertices and the decays of long lived particles such as tau leptons and heavy quarks. High hit detection efficiency and excellent spatial resolution are required for a good performance of the trajectory reconstruction, together with a fast response needed to assign each signal to the correct bunch crossing. Such demands motivate the choice of high granularity silicon sensors. While strips can be installed in the outer section, at radii between 20 and 116 cm, the high detector channels occupancy at smaller distances from the interaction region imposes the use of the pixel technology. As they operate close to the beams, the sensors and readout electronics must be sufficiently radiation hard to guarantee good performance and a long lifetime despite the exposure to heavy irradiation. The structure of the tracker detector is shown in Fig. 2.5. The various subsystems are described in the following sections.

2.3.1 The pixel detector

The silicon pixel detector is the innermost active section of CMS. It provides on average three spatial measurement points that are used as seeds to initiate the track reconstruction for charged particles and to identify primary vertices. Given the proximity to the beam collision region, the hit detection efficiency and spatial resolution achieved by the pixel detector heavily

influence the precision of the initial particle position and propagation direction measurement, crucial for the identification of primary vertices and of displaced decays. The accurate estimation of the particle momentum from the track curvature in the magnetic field requires on the other hand a bigger lever arm, and is hence mostly provided by the silicon strip layers.

The choice of using pixel sensors guarantees the required high spatial precision, allowing for hit resolutions between 9 and few tens of μm. It has nevertheless other advantages in the high particle densities typical of the region close to the beams. It limits the occupancy of the detector channels down to values as low as 10^{-4} per bunch crossing, and provides a fast response, needed for the association of the hits to the correct bunch crossing.

The pixel detector consists of two parts, the barrel (BPIX) and the forward (FPIX), as shown in Fig. 2.6. Together they provide three measurement points in almost the entire central tracker pseudorapidity range of $|\eta| < 2.5$.

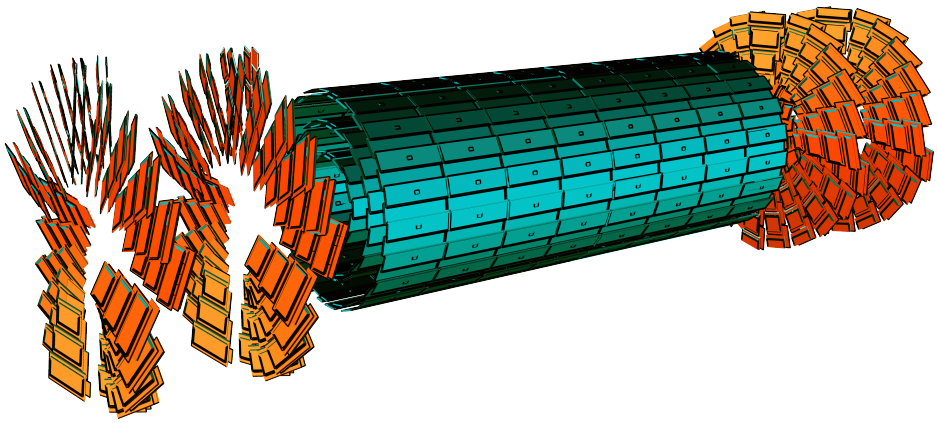

Figure 2.6: Sketch of the CMS silicon pixel detector. The barrel (BPIX) and the forward (FPIX) parts are visible in green and orange respectively.

Pixel barrel detector (BPIX)

The BPIX is made of three 53 cm long cylindrical layers at radii of 4.4, 7.3 and 10.2 cm from the beam axis, and has a modular structure. The detector units are mounted on a two-shell mechanical skeleton made of alluminum pipes of trapezoidal section, which distribute the liquid fluorocarbon (C_6F_{14}) needed to maintain the detector at the operational temperature. The structure is visible is Fig. 2.7. Thin carbon fiber ladders are glued to alternating sides of the pipes, as shown in Fig. 2.7, providing support for the modules with only a small contribution to the total material budget. Half ladders and half modules are installed at the connection between the two half-shells. The half layers contain 8, 14, and 20 ladders in the first, second

and third layer respectively.

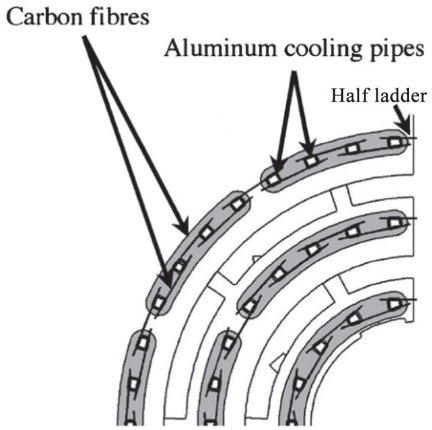

Figure 2.7: Cross-section of the BPIX detector, showing as ladders and half-ladders are mounted on the cooling pipes.

The active barrel layers end at both sides with a flange which provides mechanical support and contains electronic boards and cooling pipes. Being in front of the FPIX rings, the flanges represent a significant contribution to the amount of material within the tracking acceptance. Two supply tubes located after the endflanges connect the BPIX to the rest of the CMS detector. They hold the power cables, the cooling lines, and part of the readout electronics.

Each barrel ladder mounts eight rectangular detector units, or modules. They have a size of 66.6×26.0 mm^2, and contain a 285 µm thick segmented silicon sensor, the readout chips, a support base, consisting of two *basestrips*, and electronic components (the *High Density Interconnect* and the *Token Bit Manager*) [41, 42]. The sensors implement the *n-in-n* concept, with pixels consisting of high dose n implants on a n substrate. The pn junction is located on the backside of the sensor. The *p-spray* solution is adopted for the inter pixel isolation. The n implants are maintained at ground potential, while the p implants are at negative voltage, such that the diode is in the reverse bias configuration, and fully depleted. This creates an electric field in the entire silicon volume.

Pixels are connected to the readout chips (ROCs) by indium bump-bonds. A full module contains an 8×2 array of ROCs, each reading out a matrix of 52×80 pixels. Pixel cells have a size of 100 μm in the $r\phi$-plane, and 150 μm along the z coordinate, the minimal area being constrained by the technology employed in the readout electronics design. Thanks to the choice of a nearly square shape, the pixel detector can achieve a very good spatial resolution along the beam direction as well (see Section 3.1.1), with significant benefit to the tracking performance. Double-area pixels are installed ad the edge of each ROC to allow the readout of the peripheral region.

Charged particles traversing the sensor ionise the medium, producing pairs of negative elec-

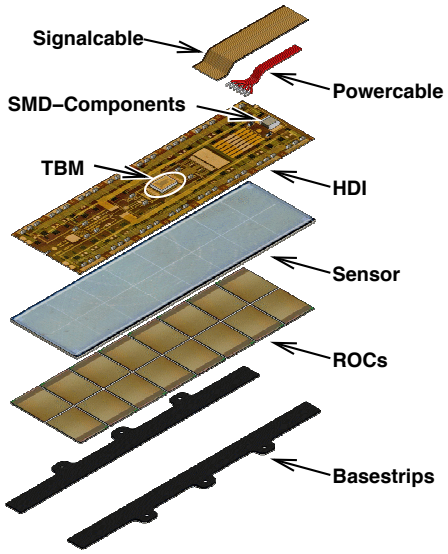

Figure 2.8: Exploded view of a BPIX detector module.

trons and positive holes. The average number of electron-hole pairs is between 25000 and 30000 for a normally incident particle, and is higher for an inclined trajectory. The charges drift along the electric field in the sensor volume: electrons to the n and holes to the p implants. Electrons from each pixel are collected and the signal is amplified and shaped by the ROC. The chip records the information about position and charge of each hit pixel, i.e. with a signal above a certain threshold, adds a time stamp corresponding to the LHC bunch-crossing, and stores such information in a memory during the Level 1 Trigger latency of 3.2 μs. The thresholds are programmed at the single pixel level, the tuning being part of the detector calibration. If the event is accepted by the trigger, a Token Bit Manager chip (TBM) installed in the module reads out all ROCs in the unit. The TBM consequently amplifies the signals, and sends them to the Analogue to Optical Hybrid (AOH) in the detector supply tube. The signals are transferred to the off-detector electronics, where they are converted to digital, and subsequently passed on to the global CMS data acquisition system, described in Section 2.7.

The pixel forward detector (FPIX)

The FPIX detector consists of two disks at each end of the BPIX. The disks extend between radii of 6 and 15 cm, and are installed at $z = \pm 34.5$ cm and $z = \pm 46.5$ cm. Each disk consists of two half-disks with the turbine-like geometry illustrated in Fig. 2.6, resulting in a 20° tilt of the detection surface with respect to the beam axis. As for the BPIX, the cooling system is integrated in the mechanical structure, to reduce the amount of passive material in the tracking volume. A half-disk consists of 12 U-shaped pipes, holding the trapezoidal beryllium panels

that provide support for the detector units. The disks are located inside the service cylinder hosting the infrastructure for mechanical support, cooling, power, and readout.

The FPIX detector implements the same modular concept as the BPIX. Each panel includes

Figure 2.9: FPIX detector panel.

four plaquettes of different sizes, with 1×2 to 2×5 ROCs, as shown in Fig. 2.9. The sensor design and the readout chain similar to the ones described in the previous paragraphs for the BPIX. The FPIX uses the *p-stop* technology for the inter-pixel isolation.

2.4 The silicon strip tracker

The silicon strip detector covers the radial region between 20 and 116 cm, and consists of three systems: the Tracker Inner Barrel (TIB) and Disks (TID), the Tracker Outer Barrel (TOB), and the two Tracker EndCaps (TEC+, TEC-). The TIB extends to a radius of 55 cm, and is composed by four cylindrical layers completed by the three TID disks at each end. The inner system is surrounded by the six TOB layers, which provide coverage in the longitudinal range $|z| < 118$ cm. The TEC+ and TEC- are installed in the region 124 cm $< |z| <$ 282 cm. Each of them consists of 9 rings, covering radii between 22.5 and 113.5 cm. The geometrical structure of the strip detector is shown in Fig. 2.5. The total active area is about 200 m^2.

As for the pixel detector, the detector architecture is modular, each module holding one (320 µm) or two (500 µm thick) silicon micro-strip sensors. Many different silicon micro-strip sensor designs are used for each of the subsystems, adapted to the particle flux, to the required spatial resolution, and to the surface that needs to be covered by instrumentation

at the corresponding radial distances from the interaction region. The inner part implements micro-strips with a pitch of 80 µm in the two innermost layers, 120 µm in the two outermost layers, and between 100 and 141 µm in the disks. Micro-strips with larger pitches, up to about 185 µm, are installed in the TOB and in the TEC detectors. The modules of the two inner layers and rings of TIB, TID, and TOB, as well as rings 1, 2 and 5 of TEC, contain two glued sensors which provide a measurement of the second coordinate (longitudinal in the barrel, radial in the disks).

The performance of the silicon strip tracker will be detailed in Section 3.2.

2.5 The calorimetry

2.5.1 Electromagnetic calorimeter (ECAL)

The CMS electromagnetic calorimeter (ECAL) [43] is a highly segmented detector composed of a barrel and two endcaps. A schematic view of the detector layout is shown in Fig. 2.10. The barrel and the endcaps cover pseudorapidity ranges of $|\eta| < 1.479$ and $1.479 < |\eta| < 3.0$ respectively. They consist of lead tungstate ($PbWO_4$) crystals, characterised by high density, short radiation length and fast response, combined with a sufficient radiation hardness. The crystal cell size in the $\eta - \phi$ plane is 0.0174×0.0174. The scintillation light produced by the shower of photons and e^+e^-pairs is collected by avalanche photodiodes in the barrel and vacuum phototriodes in the endcaps. A fine-grained lead-silicon preshower detector [44] is installed in front of the ECAL endcaps, with the purpose of distinguishing between prompt photons and neutral pion decays.

The relative energy resolution of an electromagnetic calorimeter can be parametrized as described in the formula

$$\left(\frac{\sigma}{E}^2\right)^2 = \left(\frac{S}{\sqrt{E}}^2\right)^2 + \left(\frac{N}{E}^2\right)^2 + C^2 \tag{2.2}$$

where S, N and C denote the *stochastic*, *noise*, and *constant* terms, respectively, measured in electron test beams. The relative energy resolution as a function of the electron energy is shown in Fig. 2.11.

2.5.2 Hadron calorimeter (HCAL)

The CMS hadronic calorimeter (HCAL) [45] is a sampling calorimeter of brass absorber material interspersed with layers of plastic scintillators as active detector material. It is located between the ECAL and the magnet in the radial distance range between 1.77 m and 2.95 m. A barrel part covers the region $|\eta| < 1.3$, and is complemented by two endcaps that extend the coverage up to $|\eta| = 3.0$. The minimal depth of the absorber is 5.8 interaction lengths, at

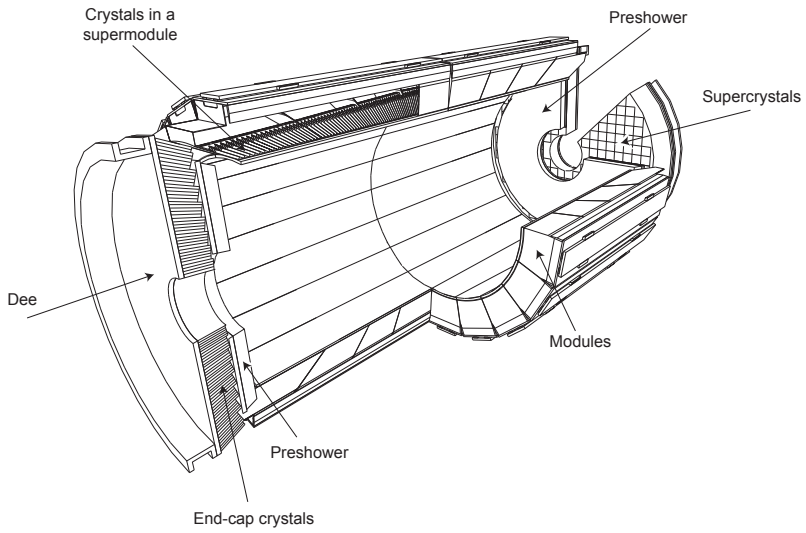

Figure 2.10: Schematic view of the ECAL subdetector. The barrel consists of two halves, each composed of 18 *supermodules*. The endcaps consist of two so-called *Dees* (named for their shapes). The preshower system is visible as well.

$\eta = 0$. The magnet and an additional layer of scintillation detectors (HO) installed outside of the coil increase the material thickness in the barrel pseudorapidity region, such that the hadronic showers are fully absorbed before reaching the muon system. This is not needed for the endcap, which has by itself a depth of at least 10.6 interaction lengths. As the ECAL, the detector is segmented. The granularity in the plane are 0.087×0.087 for $|\eta| < 1.6$ and 0.17×0.17 for $|\eta| > 1.6$.

The performance of the HCAL detector was studied using test beam data and Monte Carlo simulation. The jet energy resolution as a function of the jet energy is shown in Fig. 2.12 for different pseudorapidity ranges.

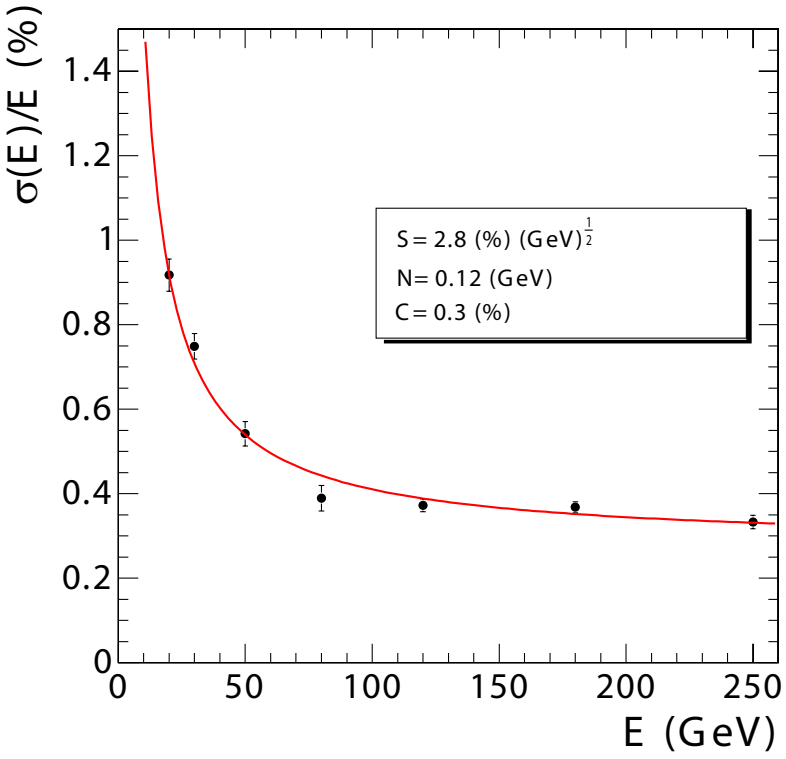

Figure 2.11: Relative energy resolution achieved by the ECAL detector as a function of the electron energy, as measured in electron test beams.

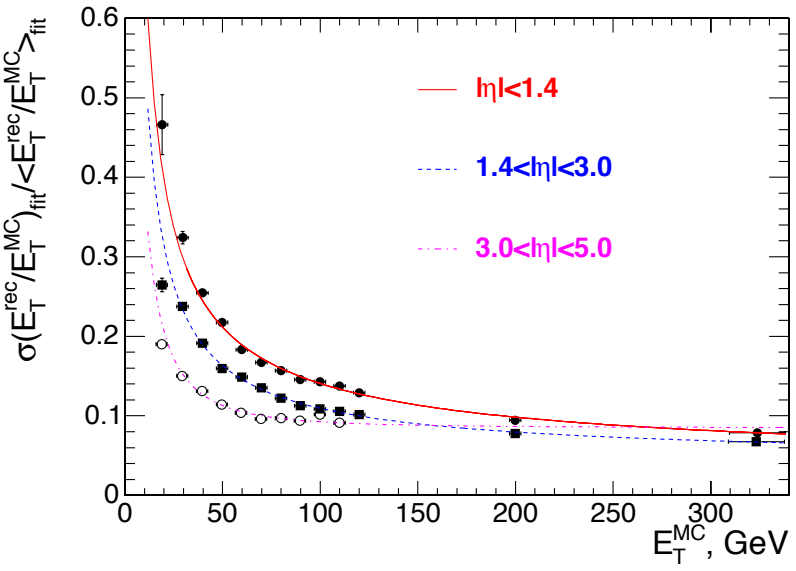

Figure 2.12: Relative jet energy resolution achieved by the HCAL detector as a function of the jet energy, as extracted from Monte Carlo simulation. Three pseudorapidity ranges are considered.

2.6 The muon system

The muon system [46] is the outermost section of CMS, hence the least exposed to the high particle fluencies typical of the LHC environment. The muon system consists of a cylindrical barrel, covering the pseudorapidity region $|\eta| < 1.2$, and two endcaps disks in the range $0.9 < |\eta| < 2.5$, implementing three different types of gaseous detectors. Four layers of detection chambers are integrated in the iron yoke, which shapes the magnetic field and provides mechanical support. A cross section of the CMS muon system is shown in Fig. 2.13

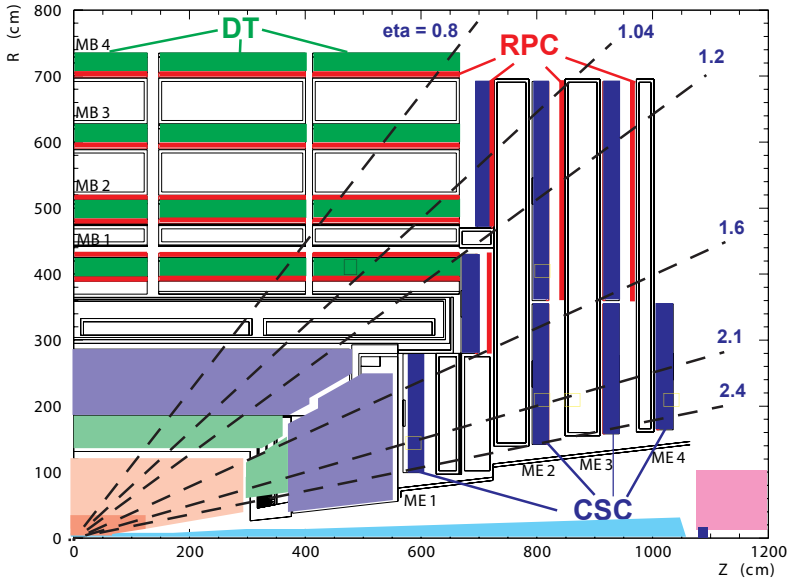

Figure 2.13: Longitudinal cross section of a quarter of the CMS muon system.

2.6.1 Drift chambers

In the barrel part, where the rate is low and the magnetic field is weaker — about 2 TeV — and uniform, chambers with standard rectangular drift tubes (DT) are used. Each of the 12 radial sectors of the five cylindrical wheels implements four stations. The three inner stations include 8 layers of tubes measuring the coordinates in the $r - \phi$ plane, and 4 layers measuring z. The outermost station contains only 8 layers for the $r - \phi$ measurement. The 2.4 m long drift cells have a transverse size of 21 mm and contain a gas mixture with 85% of Argon and 15% of CO_2, corresponding to a maximum drift time of 380 ns. The tube dimension is a compromise between the need for a low occupancy, and the exigency to limit the number of channels. The maximum drift distance is 2.0 cm. The spatial resolution of each measurement point is approximately 200 μm, resulting in a precision of 1 mrad in the determination of the

transverse muon propagation direction.

2.6.2 Cathode strip chambers

In the muon endcaps DTs are replaced by trapezoidal multiwire proportional chambers (cathode strip chambers, CSC). They can operate at higher rates and in a strong and non-uniform magnetic field, and are characterised by a fast response time. Each endcap consists of four stations of 3.4 m long and 1.5 m wide chambers, with 6 anode wire planes interleaved among 7 cathode panels. Wires are oriented azimuthally, while strips run along the radial direction. Together they provide a measurement of both $r - \phi$ and η, with a resolution of approximately 200 µm.

2.6.3 Resistive plate chambers

Thanks to their efficiency and their good background rejection, DTs and CSCs are used to trigger on the muon p_T. Nevertheless, at high instantaneous luminosity a complementary system is needed to provide accurate timing information for the association of the events to the correct bunch crossing. This consists of a set of resistive plate chambers (RPC), arranged in 6 layers in the barrel and 3 in the endcap, covering together the central pseudorapidity range $|\eta| < 1.6$.

2.7 Trigger and data acquisition system

The LHC collisions, characterised by a rate of 20 MHz, produce an enormous amount of data, such that it cannot be stored and processed. A complex trigger system analyses each bunch crossing and applies a first but rather sophisticated event selection, which reduces the rate to few hundreds of Hz, within the limits of the current data recording technology and of the available CPU power needed for the reconstruction.

The system includes two stages: the Level 1 Trigger (L1) [47] and the High Level Trigger (HLT) [48, 49]. The L1 consists of programmable electronics which decides whether to accept or reject an event based on low granularity data from the calorimeters and the muon systems. The choice of processing only data from few of the CMS subsystems is dictated by the need to reduce the decision time, or *latency*, during which the signals from the hit units have to be stored in local memory buffers. The size of these memories and the LHC bunch spacing constrain the L1 latency to be smaller than 3.2 µs. The L1 output rate is of the order of 100 kHz.

The L1 trigger accesses the information from the muon and from the ECAL systems to perform a first elementary muon and electron identification. For the electrons, it reads out the energy

deposits in 5×5 matrices of ECAL crystals — the *trigger towers* — and compares their sum to a given threshold. A rudimentary isolation requirement can also be applied, taking into account the energy deposits in neighbouring towers and in the corresponding HCAL cells. Muons provide a clean signature, and can be reconstructed with good precision even at the hardware trigger stage. Hits in the muon stations are detected by the L1 electronics, and trajectory candidates are built in the DT and CSC systems using a fast track finding algorithm. The four candidates with highest p_T and best quality are combined with the information from the RPC to build a L1 global muon object, which has to pass a given p_T threshold. As well as for electrons, additional requirements on the isolation can be applied, using the information from the calorimeters.

The High Level Trigger is a software system, implemented in a farm of thousands of processors, which further suppresses the rate to few hundreds of Hz. It starts from the L1 objects, and exploits high resolution data from all subdetectors to perform a partial reconstruction of the event in three sequential stages: Level 2, Level 2.5 and Level 3. To limit the CPU usage, the most intensive steps — such as tracker hit reconstruction and tracking — are applied only in the limited detector regions around the L1 identified candidates. Interaction vertices and high level objects — such as tau leptons and b-quark jets — are identified using simplified and faster versions of the algorithms used in the offline data processing.

The HLT software has constantly evolved and improved in the past years, and succeeded in handling the increasing data rates following the ramping of the LHC instantaneous luminosity. The online event selection has now reached an unprecedented level of complexity, unimaginable in the past generation of high energy physics experiments.

The CMS DAQ [48] is integrated between the two stages of the trigger system. When a L1 accept signal is received, it reads out the front-end electronics, combines the data into the proper event format and transmits them to the HLT farm for the second trigger selection. It forwards data to the online data quality monitoring system (DQM), which allows to monitor the quality of the data from all subdetectors in real time, and transfers the information of HLT accepted events to storage in the CERN computing facilities. The output data are in RAW format, which contains the information about the signals deposited in the detectors modules. Before being used for physics analyses, they have to be further processed by a set of software programs performing the event reconstruction, as described in Chapter 3.

2.8 Data processing

The CMS detector produces an amount of data of the order of 15 TB per day. This enormous volume of data is reconstructed, analysed and stored using a network of computing centres, the LHC computing grid, or WWCG. Computing sites are organized in a tiered structure, with the tier 0 being the CERN Computer Centre, responsible for the first (*prompt*) reconstruction

of the raw data, performed within 48 hours after the data recording. Eleven tier 1 sites are used for the event reconstruction as well, and for large scale reprocessing of data, while data analyses mostly rely on tier 2 centres. The grid infrastructure consists in total of few hundred thousand CPUs, and has a data storage capacity of about 230 PB.

2.9 Summary

This chapter gave an overview of the CERN accelerator complex and of the structure of the CMS experiment. Section 2.1 is focalised on the LHC machine and on its performance. Section 2.3 described in detail the CMS silicon tracker, while Section 2.5 and 2.6 presented the calorimeters and the muon detection system. The trigger and data acquisition were discussed in Section 2.7.

CHAPTER

THREE

EVENT RECONSTRUCTION

The data of a collision event is recorded as a set of signals containing the information about position, timing, and in some cases pulse height, proportional to the energy deposit in the sensitive material. A complex software infrastructure processes this information to build the high-level objects, such as muons, electrons or hadronic jets, used by physics analyses, in other words of performing the *event reconstruction*.

This chapter gives an overview of the reconstruction architecture and of the various reconstruction algorithms. Section 3.1 and 3.2 present the algorithms and performance of the local hit reconstruction in the pixel and silicon strip subsystems. Section 3.3 and 3.4 describe the track and vertex reconstruction. The construction of higher-level physics objects (muon, electrons, jets) is discussed in Section 3.5. Section 3.6 presents the inclusive vertex finder algorithm used for the identification of b-hadron decays in this work.

3.1 Pixel local reconstruction

The *pixel local reconstruction* is a sophisticated set of algorithms processing the raw signals collected by the pixel readout chain to estimate the hit positions and their uncertainties [50]. This algorithm provides the global tracker reconstruction with the identification number of the hit detector unit, and two coordinates specifying the hit position on the sensor surface and its error. In the BPIX, those coordinates refer to a local reference frame with the x-axis in the global $r\phi$-plane, the y-axis along the global z direction, and the z-axis perpendicular to the module surface. These local coordinates are easily converted into the global coordinates needed by the tracking algorithm.

The pixel hit reconstruction exploits the fact that in most cases the passage of a particle produces a signal above threshold in more than one cell (*charge sharing effect*), creating a set (*cluster*)

of hit pixels. For the BPIX, in the longitudinal direction, the charge sharing effect is only due to the trajectory inclination with respect to the sensor surface. In the central part of the detector, where tracks are nearly perpendicular to the modules, clusters usually consist of two or three pixels, while they can be as long as 10 pixels in the high pseudorapidity region. In the transverse plane, particle trajectories are mostly perpendicular to the sensor surface. Charge sharing in this case is ensured by the CMS magnetic field, which forces the deposited charges to drift transversely inside the sensor volume towards the neighbouring pixel cells. This phenomenon is known as Lorentz drift, and is quantified by the *Lorentz angle* variable θ_L shown in Fig. 3.1. The Lorentz angle depends on the magnetic field intensity and on the bias voltage. For an non-irradiated BPIX sensor it is approximately 23°. After irradiation, the bias voltages have to be increased to allow for an efficient collection of the deposited charges, and therefore the Lorentz angle is lower, resulting in a reduction of the charge sharing.

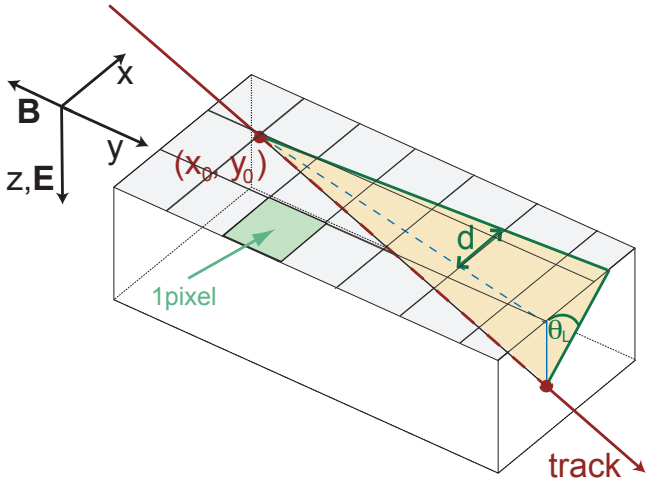

Figure 3.1: Schematic view of a BPIX sensor with the definition of the local reference frame. The charge sharing effect due to track inclination and Lorentz drift is shown. θ_L indicates the Lorentz angle.

The large majority of clusters is constituted by two hit pixels, while single-pixel clusters can be produced by trajectories with very particular inclinations, or in the cases where the signal in the second pixel is not high enough to pass the threshold. Typical BPIX cluster length distributions are shown in Fig. 3.2. In the FPIX, the charge sharing effect in the longitudinal direction is enhanced by the peculiar disk geometry, which results in a 20° inclination of the sensor surface with respect to the incoming particles. A combination of geometry and Lorentz drift ensures the effect in the radial direction.

The first step in the pixel local reconstruction is the so-called *cluster finding*. It starts by identifying a pixel cell with a high signal-to-noise ratio — the *seed* — and it searches and merges adjacent pixels that are also hit. The total cluster charge is computed and compared to a

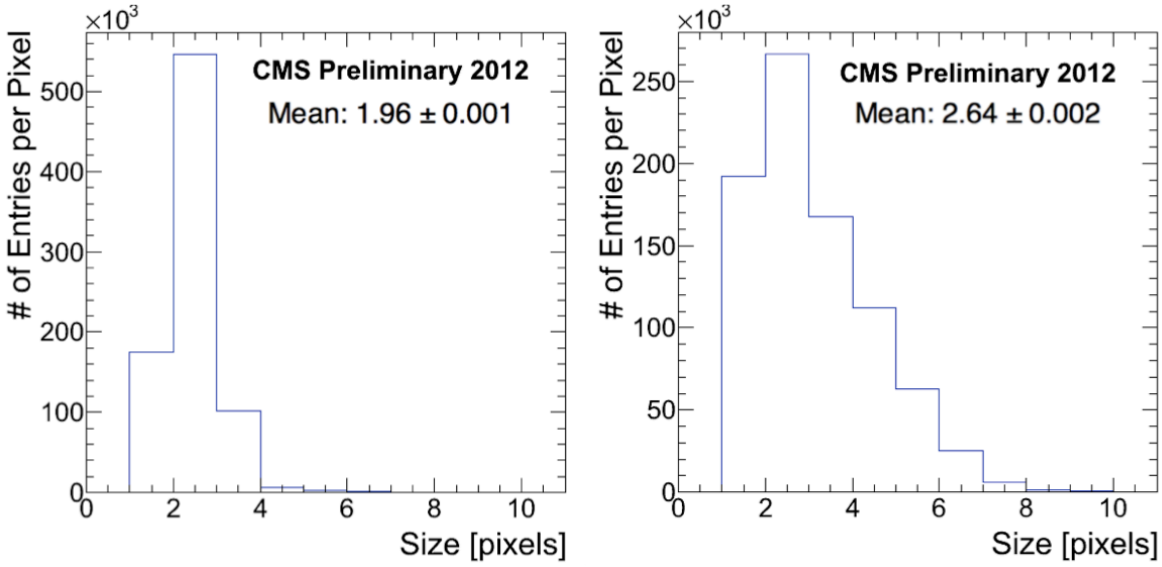

Figure 3.2: Distributions of BPIX cluster length in the transverse plane (*left*) and in the longitudinal direction (*right*), in 2012 collision data.

threshold.

Once the cluster is built, the hit position is determined. CMS employs a sophisticated algorithm, exploiting the signal heights in each pixel cell.

The hit position reconstruction consists of two iterations. The first step performs a weighted average of the charge deposit distribution, applying a correction for the Lorentz drift in the x direction. This procedure gives an approximate estimation of hit positions and errors to be used in the first iterations of track reconstruction. The second step is applied after temporary tracks are built, as it requires the information about the track impact angles. It performs a fit of the x and y projections of the charge deposit distribution in the various pixels within a cluster using predetermined shapes called *templates*. The algorithm includes also an estimation of the errors. The cluster templates are determined with a detailed simulation of the pixel sensor response implemented by the PIXELAV program [51]. It realistically models the charge deposition of tracks in the silicon medium, the electric field profile associated to the adopted sensor design, the drift of the electron-hole pairs in the material, and the effects of radiation damage. This last feature allows to adapt the templates to the increasing radiation doses [52]. The use of the template algorithm based on PIXELAV improves the transverse hit resolution by approximately 1 µm, and the longitudinal by few µm.

3.1.1 Performance of the pixel detector local reconstruction and resolution measurement

The reconstruction of charged particle trajectories is mostly seeded by triplets or doublets of hits in the pixel detector. In the absence of such hits, the tracking algorithm is still capable of detecting the particle, but the determination of the track parameters will be less precise. Hence, the local reconstruction has to guarantee a high hit detection efficiency. The efficiency is measured with proton-proton collision data collected in 2011. The result is shown in Fig. 3.3, for the three barrel layers and the four forward disks. The efficiency is higher than 99.5% in all subsystems.

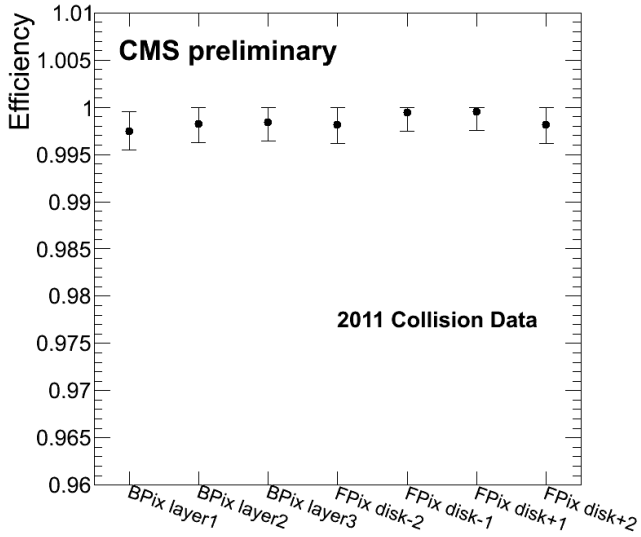

Figure 3.3: Hit detection efficiency for the three pixel barrel layers and the four endcap disks, as measured in data 2011.

For a pixel detector, a crucial performance parameter is the spatial resolution. It is determined by the pixel cell size, charge sharing between adjacent pixels, and reconstruction algorithm. The resolution is also sensitive to the operating conditions of the detector, such as bias voltages, magnetic field, signal thresholds, and can be heavily degraded by irradiation. A measurement of the hit position resolution as a function of the cluster size is relevant to test the template reconstruction procedure, and can help in tuning it. The measured position resolution should be consistent with the uncertainties assigned by the algorithm. An inaccurate estimate of the hit errors in the proximity of the primary interaction would in fact propagate to the uncertainties on most of the track parameters, and consequently to the high level tools used for the identification of long lived particle decays.

The present section describes the measurement of the spatial resolution achieved by the BPIX

detector with data collected at the beginning of the LHC data taking, in 2009 [53] and 2010. The results refer to non-irradiated sensors. The resolution values are being monitored along with the increasing integrated luminosity, in order to keep the effects of radiation damage under observation.

A standard technique to estimate the hit position resolution is based on the measurement of the *inclusive hit residuals*. For each hit, residuals are defined as the distance between the reconstructed hit position within a module and the predicted local coordinates, estimated by extrapolating the track to the sensor surface. The resolution is defined as the σ from a Gaussian

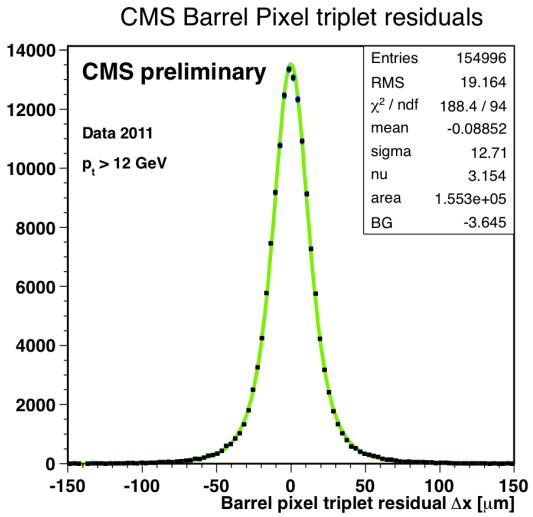

Figure 3.4: Hit residual distribution for the pixel barrel detector in the transverse plane, with 2011 collision data.

fit of the residual distribution, as shown in Fig. 3.4 for the local coordinate x. The resulting width, $\sigma_{\text{residuals}}$, includes two contributions:

$$\sigma^2_{\text{residuals}} = \sigma^2_{\text{hit}} + \sigma^2_{\text{extrapolation}}, \tag{3.1}$$

where σ_{hit} is the intrinsic resolution of the pixel detector, and $\sigma_{\text{extrapolation}}$ is the uncertainty associated to the extrapolation of the charged track.

This thesis presents an alternative technique, the so called *overlap method*, which minimizes the track extrapolation error and is sensitive to the intrinsic hit resolution. The method was developed and applied for the first time to cosmic data in 2008 [54] , and was later applied to LHC collision data.

The measurements are compared to the prediction of the standard CMS simulation, and PIXELAV.

The overlap method

The overlap method uses charged particle tracks that traverse the sensors in the overlap (*overlap sites*) between adjacent modules, leaving two hits within the same layer. Such hit pairs are selected by the proper requirement on the identification numbers of the two detector units.

The technique is illustrated in Fig. 3.5. As in the case of the inclusive residuals, the coordinates

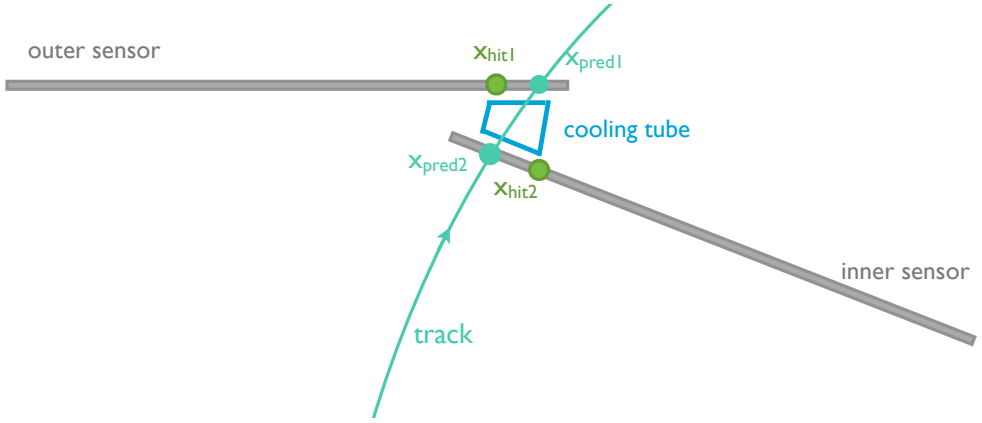

Figure 3.5: Schematic view of a pair of overlapping modules in the BPIX detector in the transverse plane, with a representation of the procedure. The distance between the two sensors is of the order of few mm.

of the reconstructed hits are compared to the predicted positions from trajectory extrapolation. To avoid a bias in the estimation of the expected hit position, the layer with the hits under study is not used in the track reconstruction. The trajectory is subsequently extrapolated inwards to get the track parameters at the surface of the outer module, and outwards to the surface of the inner one. Those parameters are then projected to the other module by the standard propagator implemented in the tracking code, and combined to obtain the best estimate of the track position and direction. As the overlapping regions of adjacent sensors are separated by a short distance — namely few millimeters — this method reduces the uncertainty from the trajectory extrapolation, which is proportional to the propagation path.
The difference Δx_{pred} (Δy_{pred}) between the two local predicted track positions within the two modules is calculated, as well as the difference Δx_{hit} (Δy_{hit}) between the reconstructed hit local positions and the double difference between them, $dd = \Delta x_{\text{hit}} - \Delta x_{\text{pred}}$ ($\Delta y_{\text{hit}} - \Delta y_{\text{pred}}$). For each overlap site the dd distribution is in first approximation a Gaussian centered in 0. The width is however evaluated using the RMS of the histogram, as the σ of a Gaussian fit would not properly account for the non-Gaussian tails.

As in the case of standard residuals, the two components from intrinsic hit resolution and track extrapolation error contribute to the dd width, although the impact of the propagation

uncertainty is reduced. This uncertainty is evaluated in each overlap site as the mean of the distribution of the errors given by the extrapolation algorithm, and is quadratically subtracted from the width, $\sigma^2_{\text{hit}} = \sigma^2_{\text{dd}} - \sigma(\Delta x_{\text{pred}})^2$, to recover the intrinsic hit resolution. This procedure gives a pair of x and y resolution measurements for each overlap site in the BPIX. An estimate of a single pair of values for the whole detector is obtained by calculating the weighted average of all measurements.

As discussed in the introduction to the present section, the pixel hit resolution depends heavily on the charge sharing effect. The dependence on the cluster size and, in the longitudinal plane, on the track inclination is therefore investigated as well.

Data samples and selection cuts

A first global estimation of the hit resolution was performed with the first collision data collected by CMS in 2009, at center-of-mass energies $\sqrt{s} = 900$ GeV and $\sqrt{s} = 2.36$ TeV. A basic selection was applied to retain events with a detected proton-proton hard scattering at the beam crossing (the so-called *minimum bias* events) while rejecting the background produced by beam interactions with the residual gas in the beam pipe (beam halo). The selection is based on the trigger information from the Beam Scintillation Counter system (BSC), consisting of two scintillators installed at a distance of 11 m from the beam interaction region at both $+z$ (BSC+) and $-z$ (BSC-). The requirement of two coincidental hits in BSC+ and BSC- is sufficient to exclude beam halo events.

In 2010 the measurement was reproduced and expanded using a significantly bigger sample of collision events at $\sqrt{s} = 7$ TeV. A trigger requirement based on the presence of high transverse momentum jets was chosen in this context.

The subsequent selection is substantially common to both measurements. Tracks are retained if:

- The total number of hits in the tracker detector is larger than 5;
- The track χ^2 probability is higher than 10^{-3};
- The uncertainty on the predicted position from trajectory extrapolation, $\sigma(\Delta x_{\text{pred}})$ ($\sigma(\Delta y_{\text{pred}})$), is smaller than 25 µm, to assure sensitivity to the intrinsic hit resolution;
- The distance between the two selected hits along the track is smaller than 2 cm, to reduce the extrapolation error from multiple scattering, which is proportional to the thickness of the material traversed by the particle;
- The track momentum is bigger than 2.5 GeV. This threshold can be increased to 5 GeV in the 2010 analysis;
- The number of crossing tracks per overlap site is bigger than 20 (30 in the 2010 analysis).

Sources of uncertainty

The main source of systematic uncertainties in the overlap method is the determination of the predicted hit positions via track extrapolation. Although the corresponding uncertainty will not be exactly quantified in this thesis, a qualitative description of the effects on the measurement is described in the present section.

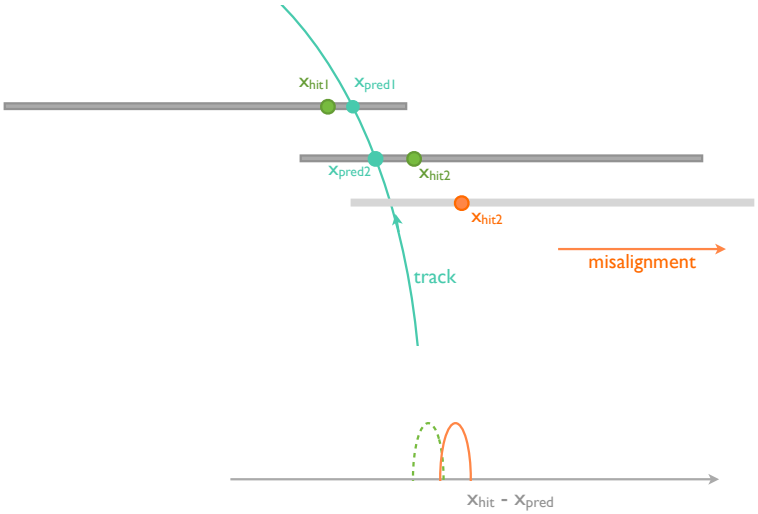

Figure 3.6: Effect of the residual translational misalignment on the resolution measurement. Translational misalignment shifts the double difference distribution, while shape and width remain unaffected.

Two main contributions participate to the total track propagation uncertainty: the first is due to detector misalignment, the second is related to the multiple scattering (MS):

$$\sigma^2_{\text{extrapolation}} = \sigma^2_{\text{misalignment}} + \sigma^2_{\text{MS}}. \tag{3.2}$$

The multiple scattering σ_{MS} depends on charge, velocity and momentum of the particle — motivating the cut on the track p_T — and on the thickness of the traversed material. It is therefore intrinsically reduced by the use of the overlap method, which requires the extrapolation of charged particle trajectories over the short distance between the two overlapping sensors. On the other hand, the amount of detector material which has to be traversed by a particle from one sensor to the other is not negligible, as the cooling pipes are located exactly in the overlap region. However, the resulting uncertainty is automatically included in $\sigma(\Delta x_{\text{pred}})$ and subtracted from the *dd* width together with it.

The $\sigma_{\text{misalignment}}$ term is due to the residual misalignment of the modules in the overlapping pair, after the application of the alignment procedure. Misalignment can be *translational*, when the estimated position of the detector units is shifted from its true location, or *rotational*. Additional effects are introduced by a mismodeling of the sensor geometry.

An advantage of the overlap method is that the measurement is not influenced by translational misalignment. As shown in Fig. 3.6, it only produces a shift of the mean of the *dd* distribution, while shape and width, from which the resolution is estimated, are not affected. An analysis of this shift could provide an estimate of the misalignment itself.

Rotational residual misalignment instead influences the measurement, as it introduces a dependence of *dd* on the track inclination, resulting in an artificial increase of the width. Based on independent estimations performed by the alignment group [55], this effect is of the order of few µm.

The last source of systematic uncertainty related to misalignment is the simplified geometrical description of the detector units implemented in the CMS software, used by the event reconstruction and reproduced by the simulation, which assumes their surface to be flat. Due to the assembly procedure, silicon sensors are instead bowed, the distance between the actual center of the sensor and the flat approximation being about 30 µm. The effect on the measurement, similar to the one caused by residual rotational misalignment, can be seen in Fig. 3.7, showing the *dd* distribution as a function of the longitudinal hit position along the sensor, for a single module pair located in BPIX layer 3. The bias from the artificial inflation of *dd* propagates to the estimation of the intrinsic hit resolution, and has to be corrected for. A more realistic model of the shape of the detector units, taking into account their curved surface, was implemented in the CMS software shortly after the conclusion of this analysis [56]. For the purpose of a resolution measurement, it is sufficient to reduce the bias by applying an additional cut that retains only pairs of hits detected in the central part of the sensor and rejects the module periphery, in particular 1 cm at the edge of the sensor. The requirement is not applied to the 2009 data, due to the limited size of the sample.

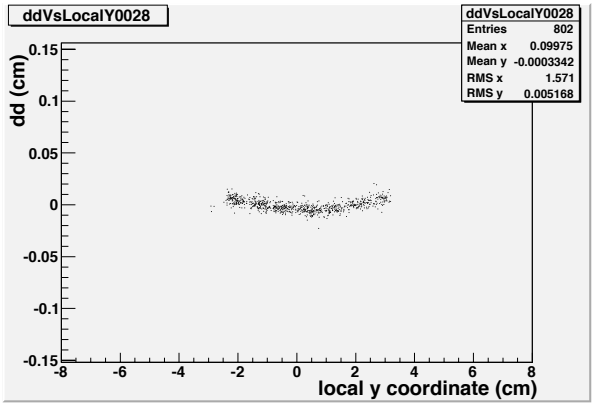

Figure 3.7: Double difference *dd* as a function of the hit position along the *y* axis. The two-dimensional distribution reflects the combination of the curvatures of the two overlapping sensors.

Results and observations

In 2009 data, only 14 overlap sites over a total of few thousands pass the selection in the entire BPIX detector, for a total of approximately 400 hit pairs. The measured transverse and longitudinal resolution values are 12.8 $\pm$ 0.9 µm and 32.4 $\pm$ 1.4 µm respectively. The exclusion of the double-area pixels installed at the edge of each ROC leads to values of 12.7 $\pm$2.3 µm along x and 28.1 $\pm$1.9 µm along the y direction. Although a differential measurement of the resolution as function of the cluster length would be desirable, the limited statistics of the 2009 run does not allow for such a detailed analysis. The results are compared to the prediction by the PIXELAV program. A sample of charged particles with the same momentum and inclination as for the tracks collected in data is generated, and consequently processed with the simulation of the sensor response. This allows to obtain a sample of hits that should accurately reproduce the data. The predicted values extracted through this procedure are 14.1 $\pm$0.5 µm along x and 24.1 $\pm$0.5 µm along y. Although data are substantially compatible with the expectation within the statistical uncertainties, a 2 σ deviation is observed. This discrepancy is due to the fact that the program does not exactly reproduce the cluster size mixture observed in the data sample, most probably because of a wrong tuning of the pixel threshold parameters in the simulation.

The larger size of the data sample recorded in 2010 allows to refine and extend this first estimation. 1316 overlap sites are found with a sufficiently high number of good crossing tracks, covering the entire BPIX pseudorapidity range. The total number of selected hit pairs is 8.3 millions. The overall results are 11.2 $\pm$0.1 µm along x and 26.8 $\pm$0.1 µm along y. The discrepancy in the comparison between the 2009 and 2010 estimations is justified mostly by a different cluster size composition of the hit samples, due to an increased presence of tracks with high pseudorapidity in the larger 2010 dataset.

The 2010 analysis investigates also the dependence of the spatial resolution on the pixel cluster length induced by the track inclination. The sample is divided into two independent sets of subsamples according to the cluster size along the x and y directions respectively. The method is consequently applied to extract a value for each size. Results are shown in Fig. 3.8, with the associated statistical uncertainties. As expected, the charge sharing improves the performance of the local reconstruction, the single-pixel hit resolution being the worst in both x and y. Along the longitudinal direction, the resolution is approximately $(pixel\ size)/\sqrt{12} \approx 43$ µm for single-pixel clusters. It reaches its best value, lower than 20 µm, for hits consisting of two pixel columns, and it slowly increases again for longer clusters. In the transverse plane, instead, although it is still worse than for hits with charge sharing, the single-pixel hit resolution is significantly better than the $(pixel\ size)/\sqrt{12}$ naive estimation, which would lead to values of about 30 µm. This is justified by the fact that the effective pixel size is reduced by the Lorentz effect, that spreads charge over an extended region, covering two adjacent pixel cells. The precision with which the hit position is known is determined by the effective pixel size, rather than by the total pixel cell size, and can therefore reach values as low as 12 µm.

The resolution values measured in data are compared to the predictions by the standard CMS detector simulation and by PIXELAV, as shown in Fig. 3.8. For a consistent comparison with PIXELAV, which does not implement the double-size pixels installed at the edge of the sensor, hits containing at least one double-size cell are excluded. As expected, this cuts mostly affects the longitudinal resolution value for the single-pixel hits, reduced from 43 to 35 µm. After some tuning of the pixel threshold parameter, PIXELAV reproduces the observed resolution within 1 µm along both the x and y directions, for all cluster sizes. Since the PIXELAV program

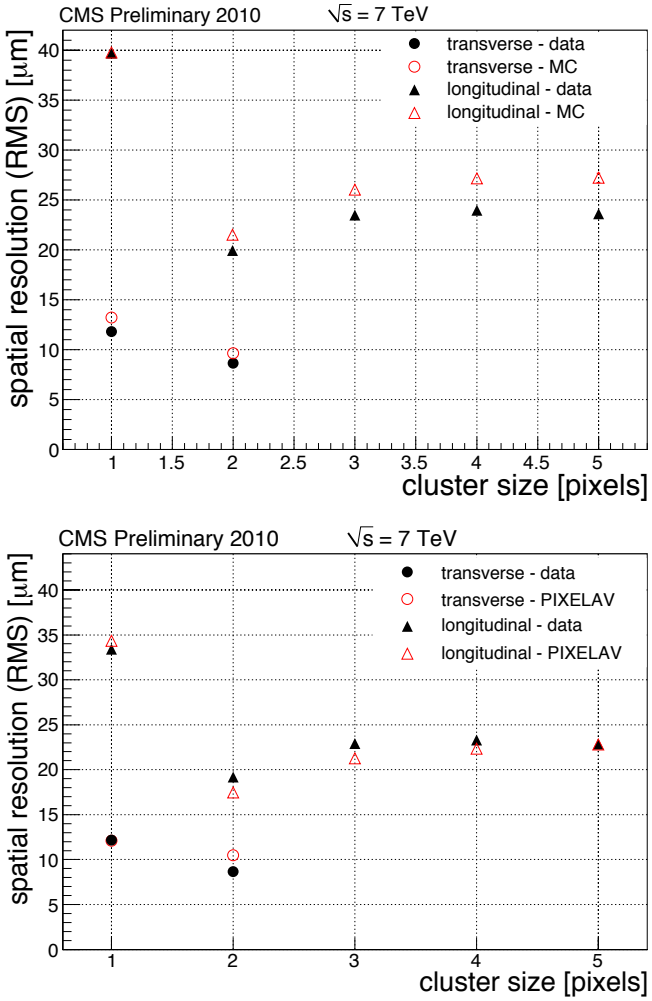

Figure 3.8: Transverse (circles) and longitudinal (triangles) hit position resolution as a function of the cluster length expressed in pixels. Only statistical uncertainties are shown. The observed numbers are compared to the predictions by the standard CMS Monte Carlo (*top*), and by the PIXELAV simulation (*bottom*). PIXELAV reproduces very well the data for all cluster size categories.

is used to produce the cluster templates used in the pixel local reconstruction, this excellent agreement is of great relevance.

3.2 Silicon strip local reconstruction

The signature of a charged particle traversing the silicon strip tracker is similar to that in the pixel detector. Due to the intense magnetic field, the deposited charges drift transversely in the sensor material, inducing a signal in a set of adjacent cells (*digis*). Digis are produced in the online reconstruction, by algorithms that run at the front end electronics. At this stage, a strip is accepted if its charge exceeds five times the expected channel noise, or if both the strip and one of its neighbours have a charge exceeding twice the channel noise. The offline local reconstruction consequently identifies sets of adjacent digis compatible with the passage of a single particle in the detector unit, and aggregates them into a cluster, providing for the seed of the final hit. Clusters are retained if their total charge is higher than five times the sum of the noises of all strips in the set. The hit position is finally determined as the charge-weighted average of all the strip positions, corrected for the Lorentz angle. The uncertainty is estimated with a quadratic function of the cluster width projected on the sensor in the plane perpendicular to the strip modules.

As for the pixel, the achieved spatial resolution depends on the particle incident angle with respect to the module surface, or equivalently on the cluster length, and on the strip pitch. It ranges between 15 and 40 µm for the different subdetectors. The values are shown in Fig. 3.9 (left). The measurement was performed with 2010 data for the barrel only (TIB and TOB), using the overlap method documented in section 3.1. Fig. 3.9 (right) shows the hit detection efficiency for all subdetectors.

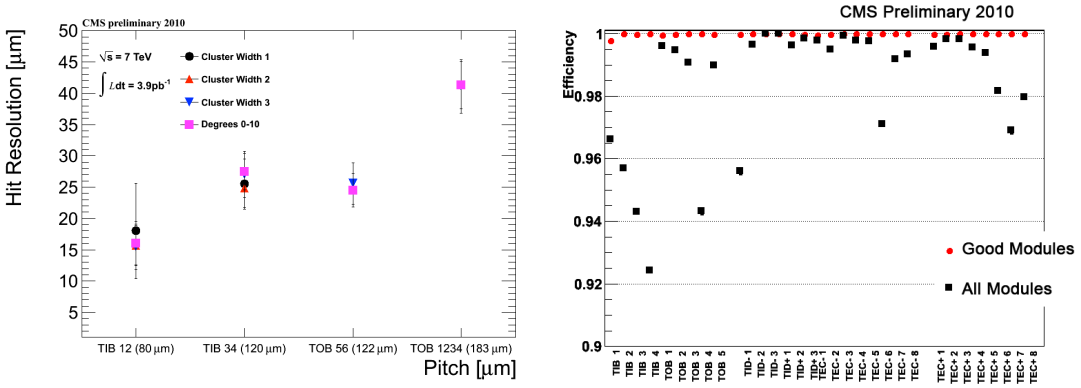

Figure 3.9: *Left*: hit position resolution for the TIB and TOB subdetectors as a function of the cluster length in µm and of the strip pitch. *Right*: hit reconstruction efficiency for all subsystems. The red points are obtained by excluding known malfunctioning detector units. Both measurements are performed using 2010 collision data.

3.3 Track reconstruction

The trajectory of a charged particle propagating in the quasi-uniform magnetic field is approximated by a helix. Five parameters are used to identify it, all defined at the impact point, i.e. the point of closest approach of the particle to the beam axis: the impact parameter — the distance between the impact point and the beamspot — in the transverse plane (d_0) and in the longitudinal (z_0) direction, the polar (ϕ_0) and azimuthal (expressed as $\cot\theta_0$) angles, and the transverse momentum p_T. The purpose of track reconstruction [53,57] is to determine these parameters and their uncertainties for each of the charged particles traversing the tracker detector, using the set of measured hits in the pixel and strip systems. CMS applies an *iterative tracking* procedure, consisting of a sequence of six iterations of the algorithm known as *combinatorial track finder*. Each iteration includes four steps:

1. The *seed generation* selects doublets or triplets of tracker hits and builds the first track candidates, providing an approximate estimate of their parameters and uncertainty. For the first iterations, only the pixel detector is scanned in the search for seeding hits. This choice is motivated by its finer granularity and its better spatial resolution compared to the outer tracker, which allows to effectively constrain the initial trajectory parameters. Moreover, it guarantees a higher efficiency, since a fraction of the charged particles produced in collisions, especially at low momentum, suffers destructive interactions or lose a significant fraction of their energy in the tracker material before reaching the outer layers. In addition to the hits, a loose beam crossing position (*beamspot*) constraint is used at this stage, in order to limit the number of possible hit combinations. In the following

iterations the seed generation moves outwards, including first the inner TIB, TID and TEC layers, and finally the two inner layers of the TOB, allowing for the reconstruction of trajectories produced in secondary interactions or decays, or without hits in the pixel detector;

2. The *pattern recognition* extrapolates the seed trajectories outwards along their expected path, searching for hits compatible with the track hypothesis. A Kalman filter is used for this purpose [58]. The trajectories are propagated first with simple analytical calculations, assuming a uniform magnetic field and neglecting the particle interaction with the detector material. The extrapolation is subsequently refined using a detailed map of the magnetic field, and accounting for possible deviations due to multiple scattering effects in matter. The algorithm identifies the detector units crossed by the charged particle, checks for possible compatible hits, and assigns them to the track candidate. The trajectory is then refitted. Once that the end of the detector is reached, the procedure is repeated inwards, from the outer layers towards the primary interaction;

3. The *track fitting* uses a Kalman filter as well. It provides the final estimate of the parameters, by refitting the trajectory to eliminate possible biases from the constraints applied at the seeding stage. The Kalman filter is initialized with the seed parameters at the innermost hit, and proceeds iteratively through the full set of hits identified by the track finding, re-estimating their position using the updated track inclination. The trajectory is then smoothed by a second fit run inwards towards the beam line;

4. The *track selection* has the purpose of rejecting tracks that are not associated to any charged particle, or *fake* tracks. It requires a minimum number of associated hits, a maximum number of detector layers without any hit, a maximum value of the fit normalized χ^2, and a minimal significance of the parameters. Selected tracks are then sorted into different collections, according to the same quality criteria, the *High Purity* being the one used for the construction of physics objects.

As mentioned above, the procedure is applied six times, with different configurations of seed generation and final track selection. The first iteration reconstructs most of the trajectories corresponding to relatively high momentum particles — typically with $p_T > 0.8$ GeV — originated close to the beamspot. The following iterations allow to recover particles with lower momentum, or produced far from the beam interaction region.

3.3.1 Track reconstruction performance

The performance of the tracking algorithm influences the capability of reconstructing high level objects, in particular primary interaction vertices and displaced decays of long-lived particles. It can be described in terms of resolution of the track parameters, track finding efficiency

and fake track rate. The set of selection cuts forming the High Purity working point of the algorithm allows to suppress the fake rate to less than 1‰.

The efficiency, defined as the fraction of charged particles that can be associated with a reconstructed track, is shown as a function of p_T in Fig. 3.10. It is obtained using a Monte Carlo sample of top pair events with additional proton hard scattering interactions reproducing approximately the pile up distribution observed in 2011 data [57]. Despite the high particle density produced by high luminosity collisions, the excellent performance of the tracker detector and the sophisticated reconstruction algorithm provide an efficiency higher than 90% in a wide region of the spectrum.

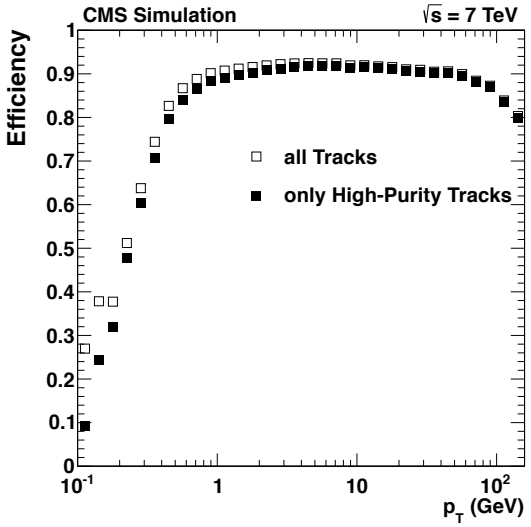

Figure 3.10: Efficiency of the track reconstruction algorithm as a function of the particle p_T. It is obtained using a sample of simulated top pair events with an average number of simultaneous p-p collisions equal to 8.

The resolution of the five track parameters as a function of p_T is described in Fig. 3.11, for the same simulated sample. Two effects contribute to determine the resolution: the multiple scattering of the charged particles in matter, and the precision of the hit positions measurement. As mentioned in Section 3.1, the multiple scattering effect is proportional to the inverse of the particle momentum. It is therefore responsible of the worse resolution values observed in the low p_T region, and drives the quick fall of the curve with the increasing p_T. For the spatial parameters, the resolution above approximately 30 GeV is approximately constant. The asymptotic value is determined by the intrinsic spatial resolution achieved by the detector, mostly by the first layer of the pixel system, the closest to the trajectory origin. A rise is observed on the other hand in the p_T resolution. This is due to the dependence of the trajectory curvature on $1/p_T$, which makes the momentum measurement for high p_T particles less precise.

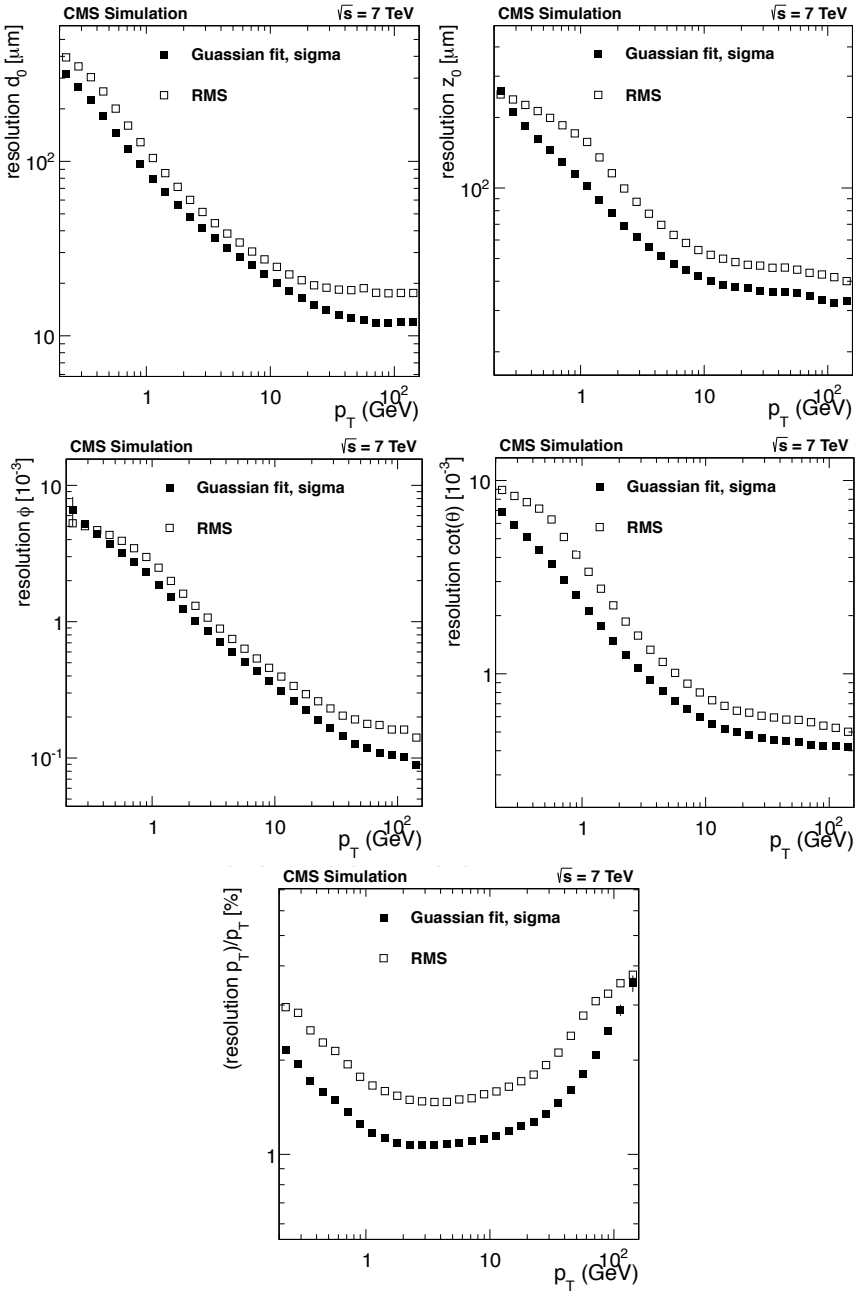

Figure 3.11: Resolution of the five parameters defining a track, as functions of p_T: the transverse and longitudinal impact parameters, the angles ϕ and $\cot\theta$, and p_T. All distributions are obtained using a simulated sample of top pair events with an average number of simultaneous p-p collisions equal to 8.

3.4 Primary vertex reconstruction

The purpose of primary vertex (PV) reconstruction is to determine position and uncertainty of each of the proton-proton interactions in a beam crossing, starting from the reconstructed tracks [57]. It consists of three steps:

1. The *track selection* stage identifies a sample of tracks originated from the beam interaction region, to be included in the primary vertex reconstruction. Requirements are imposed on the transverse impact parameter significance with respect to the beamspot, the total number of tracker hits and the normalised χ^2. No p_T threshold is applied;

2. Selected tracks are clustered into a set of vertex prototypes by a *vertex finding* algorithm, according to the z position of their point of closest approach to the beamspot. The x and y coordinates are not used, given the small transverse size of the beam interaction region. The resolving power needed to identify up to several tens of primary interactions in the high luminosity LHC runs is guaranteed by the sophisticated *deterministic annealing* algorithm (DA) [59]. It builds a χ^2 function of the configuration of the vertex positions and of the track-to-vertex assignment probabilities (weights). The χ^2 is minimized through a complex iterative procedure analogous to the gradual cooling of a thermodynamical system in statistical mechanics. The result is a sample of vertex candidates, each of them associated to a set of tracks with weight greater than 0.5. The DA algorithm is robust against outlier tracks and misassignments, and provides a resolving power of about 1 mm;

3. The last step of vertex reconstruction is the *vertex fitting*, which determines position and uncertainty of each candidate from the corresponding set of assigned tracks. Of the several methods developed in CMS through the years, the most robust is the so-called *adaptive vertex fitter* (AVF) [60], consisting of an iterative weighted Kalman filter. As in the case of the DA clustering, a weight w_i is assigned to each track i as a function of its compatibility with the vertex prototype, measured by a χ^2 variable, as in

$$w_i(\chi^2) = \frac{exp(-\chi^2/2T)}{exp(-\chi^2/2T) + exp(-\chi^2_c/2T)} \tag{3.3}$$

   The constant χ^2_c defines the threshold at which the weight is 0.5: above this threshold, a track is consider an outlier and is rejected. The weight curve shape as a function of χ^2 is controlled by a *temperature* parameter T. At high temperature, the weight is smooth and almost constant, while at $T = 1$ it becomes equivalent to a sharp cutoff at χ^2_c. In the fitting procedure, the temperature is set to an initial value $T_{in} > 1$, and the weighted sum of the squared distances of all tracks from the vertex position is minimized as prescribed by the Kalman filter method. As in the DA clustering, an *annealing* procedure is then applied: the temperature is gradually lowered in a sequence converging to 1, and the

Kalman filter is run again at each step until $T = 1$ or until the vertex position is stable within 1 µm. This technique allows to avoid local minima and to stop the iteration only when the absolute minimization is reached.

The AVF algorithm is robust and adaptable to several applications besides the reconstruction of primary interaction vertices. With the appropriate tuning of the parameters, it is used for the identification of displaced vertices from long-lived particle decays, such as b hadrons, and for kinematic fitting.

3.4.1 Vertex reconstruction performance

The relevant parameters defining the performance of vertex reconstruction are the spatial resolution, the efficiency and the fake vertex rate.

The PV resolution is of great relevance as it affects the identification of multiple proton-proton interactions at large instantaneous luminosities, and propagates to the uncertainty on the reconstruction of displaced decays. The resolution depends on the number of tracks used in the vertex fit, and on the intrinsic uncertainty on the single track spatial parameters, which is mostly dependent on the track p_T. The results of a measurement performed with data are shown in Fig. 3.12. The resolution is above 100 µm for vertices consisting of few tracks, and

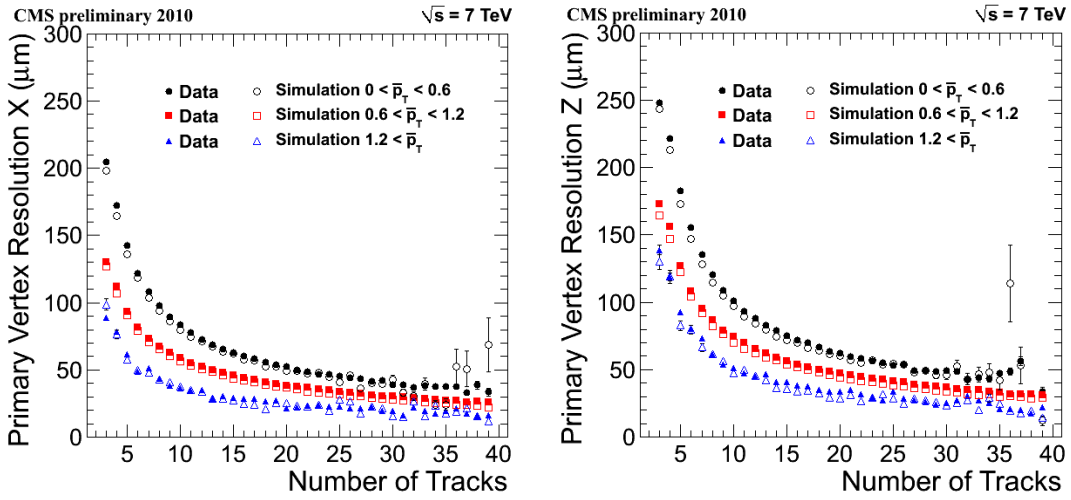

Figure 3.12: Primary vertex position resolution in the transverse (*left*) and longitudinal (*right*) planes, as a function of the number of tracks included in the vertex fit, and for track samples with different average p_T.

decreases with $1/N_{tracks}$ down to asymptotic values of approximately 20 µm for $p_T > 1.2$ GeV. Although the described study is aimed at measuring the resolution of the proton-proton pri-

mary interaction vertices, the conclusions are not specific, and can be applied to the identification of secondary vertices from long-lived particle decays, as the same algorithms are used in the two contexts.

As the resolution, the PV reconstruction efficiency depends on the number of tracks forming the cluster, and on their average p_T. Efficiency is measured in data, and is found to be between 98%, for clusters of two or three tracks, and 100%. The effect of high pileup is estimated to be negligible. The fake vertex rate is strongly suppressed by the use of the DA clustering algorithm, down to approximately 1%.

3.5 Physics objects

The information from tracking, calorimeters and muon system is combined in the construction of the high level objects to be used in data analysis: muons and electrons, tau leptons, photons, and hadronic jets. Tools are available to identify heavy-flavour particles — hadrons containing a charm or a b quark — within a jet, based on their heavy mass and long lifetime.

3.5.1 Leptons

The local reconstruction within the muon system — the *standalone muon reconstruction* — is similar to the identification of tracks in the silicon tracker [61]. In a first step, the sequence of hits left by the traversing muon in each single muon chamber, DT, CSC or RPC, is linearly fit into a track segment, which provides a preliminary estimation of the muon propagation direction and momentum. All the segments and the corresponding hits are subsequently fitted together into a single standalone track using a Kalman filter technique. In the reconstruction of collision data, a constraint on the beamspot position is applied to improve the momentum resolution.

A muon produced in LHC collisions gives both a standalone muon candidate, reconstructed in the external chambers, and a track in the silicon tracker. The tracker dominates the momentum resolution up to about 200 GeV, while a precise measurement of the small curvature of trajectories with higher p_T requires a bigger lever arm. Therefore, to optimise the resolution, the tracker track and the standalone muon are combined together into a single global muon object. Two approaches can be adopted for this purpose: the outside-in *global muon* and the inside-out *tracker muon reconstruction*. The global muon identification starts from the muon chambers. For each standalone muon the best matching track is identified, and a Kalman filter refits the entire sequence of hits in the muon chambers and in the silicon tracker into a single trajectory, taking into account magnetic field, energy losses, and multiple scattering. The tracker muon reconstruction starts instead from the silicon tracker. All tracks with p_T above a certain threshold are considered as potential muon candidates, and are extrapolated

towards the muon subdetector. If at least one standalone muon segment matches the trajectory, a tracker muon is built and the set of tracker and muon hits are refitted into a global object. All muon candidates are finally merged into a single collection.

The characteristic signature of an electron traversing the CMS detector is its electromagnetic shower in the ECAL. Due to the fine granularity of the calorimeter, and to the strong magnetic field, showers tend to spread in the transverse plane over few adjacent crystals. Hence, electron reconstruction [62] starts by identifying such groups of hit crystals, and by merging them into a so-called ECAL *supercluster*, providing for the seed of the electron candidate. The supercluster is subsequently matched to a track seed in the silicon tracker, and a trajectory is built using a Gaussian Sum Filter (GSF) technique, which implements an accurate modeling of the electron energy loss in the detector material [63].

3.5.2 Global event reconstruction: the particle flow technique

Particle flow is one of the most sophisticated event reconstruction techniques currently applied at high energy physics experiments [64]. It combines the information from all CMS subsystems towards a reliable identification of all types of stable particles produced in LHC collisions and a precise measurement of their momentum and energy. The list of particles provided by the particle flow is used in physics analyses to build high-level objects such as hadronic jets and missing transverse energy, to identify tau leptons, and to isolate the products of the relevant primary interaction from other tracks from simultaneous pp collisions.

The elements constituting the input to the particle flow are charged particle tracks reconstructed in the tracker, calorimeter clusters, and muon tracks. A link algorithm extrapolates the elements beyond the boundaries of the corresponding subdetector, and connects them into *blocks* according to geometrical distance between them in the case of tracker tracks and ECAL and HCAL clusters, or to the χ^2 of the global fit for tracker and muon track segments. The blocks are subsequently processed by the particle flow algorithm, that identifies the type of particle in a series of sequential steps:

1. Particle-flow muons are built first, by combining the global muon object described in Section 3.5.1 with the energy deposited in the calorimeter. The corresponding tracker and muon tracks are removed from the respective collections;

2. The following step consists in the construction of particle-flow electrons. The algorithm selects tracker tracks characterised by a significant energy loss through the silicon layers, which have a high probability of being produced by an electron emitting Bremsstrahlung radiation. A combination of tracker and ECAL variables is used for the final identification. As for muons, the track segments and ECAL clusters corresponding to a reconstructed electron are excluded from the following steps;

3. Charged hadrons are built as well by connecting the tracker tracks to the ECAL and HCAL clusters. Some conditions are imposed on the compatibility between the track momentum and the calorimetric energy deposits;

4. The ECAL and HCAL clusters remaining after all the particle-flow steps, which cannot be linked to any tracker track, are identified as photons or neutral hadrons.

At high pileup, particle-flow reconstruction includes an additional step, consisting of the association of each particle-flow particle to the corresponding primary interaction vertex. This matching, performed by extrapolating the particle propagation direction towards the beam line, allows to isolate the products of the proton-proton interaction that is considered physically relevant, while rejecting objects from the additional activity in the event.

3.5.3 Jet reconstruction

A hadronic jet is a collimated stream of charged and neutral hadrons, resulting from the hadronisation of a parton — quark or gluon — produced in the hard scattering. Jet reconstruction is responsible of combining together hadrons coming from the same hard parton, providing for the link between the individual particles, experimentally observable, and the products of the hard process. A jet reconstruction algorithm accepts as input simple tracks or energy deposits in the calorimeters, as well as the high-level objects constructed by the particle flow, or generated particles in the Monte Carlo simulation.

At CMS, jet reconstruction is based on the *anti-* k_T algorithm [20]. It iteratively clusters particles according to their distance in azimuthal angle and rapidity, and to their momentum, within a cone of arbitrary ΔR radius. The cone size, set to 0.5 at CMS, intrinsically prevents the experimental access to the collinear production of parton pairs. Quarks or gluons with propagation directions separated by $\Delta R < 0.7$-0.8 cannot be resolved, and are therefore generally reconstructed as a single jet. Studies are ongoing at the time of this thesis, aimed at developing techniques to disentangle two partons hadronising into a single wide jet, based on the spatial density of particles within the jet and their momentum (the so-called *jet substructure*).

3.5.4 flavour identification: b-jet tagging

The measurement of a physics process often requires to determine the flavour of the partons produced in the hard scattering, in particular to identify charm and b-quarks. This is the purpose of the techniques known as *b-tagging*. Several algorithms are available at CMS, most of them based on reconstructed jets, exploiting the relatively large mass and long lifetime of heavy flavour hadrons, and in some cases their semileptonic decays. In general, the output consists of a *discriminator*, a variable that is related to the probability of a jet being a b-quark (or charm) jet. The most widely used discriminators in CMS are listed below:

- In the *Track Counting* (TC) approach, a jet is tagged as b-jet if at least N tracks have an impact parameter significance exceeding a certain value S, where N and S are configurable parameters defining the algorithm working point;
- The *Jet Probability* (JP) algorithm evaluates the displacement of all tracks with respect to the primary vertex. The discriminator consists of a probability density function, built combining the probability values assigned to all the tracks under the hypothesis of being originated in the PV;
- Semileptonic decays of b hadrons are used in the p_T^{rel} approach, which exploits the lepton transverse momentum with respect to the jet axis. This variable is larger for leptons produced in b-hadron decays, due to the heavy mass of the mother particle;
- The *Simple Secondary Vertex* (SSV) method is based on the reconstruction of a displaced secondary vertex within the jet, its discriminator being a monotonic function of the vertex three-dimensional flight distance. As for the TC method, different variants of the algorithm are available: the high efficiency (SSVHE), requiring at least two tracks associated to the secondary vertex, and the high purity (SSVHP), requiring a track multiplicity larger than three. As a consistent fraction of B-hadron decays lead to three charged daughters, a tighter requirement on the track multiplicity allows to enhance the b-quark component while reducing the contamination from charm and light backgrounds;
- The *Combined Secondary Vertex* (CSV) is a more sophisticated version of the SSV. It combines the reconstruction of displaced vertices with additional lifetime-related information, allowing for the identification of a b-jet even in the absence of a secondary vertex. Many variables — the vertex 2D flight distance significance, mass and track multiplicity, the fraction of energy carried by the tracks associated to the vertex compared to the total jet energy, total number of tracks in the jet, and 3D impact parameter significance of all tracks in the jet — are evaluated and included in a likelihood function used for the discrimination of light, charm and b-jets.

The performance of b-tagging algorithms is evaluated in terms of the efficiency for correctly identifying a true b-jet, and the mis-tag rate. Fig. 3.13 shows these properties for two categories of jets, as extracted from the Monte Carlo simulation.

Besides the jet-based techniques listed above, another method, the *inclusive vertex finder* (IVF), has been developed, which purely relies on track and secondary vertex reconstruction. A detailed description of the IVF is given in Section 3.6.

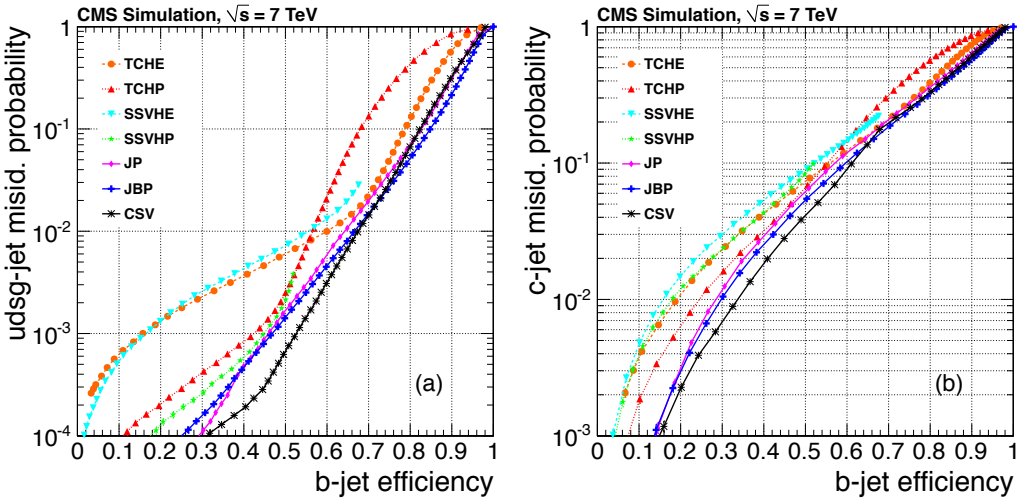

Figure 3.13: Efficiency for identifying a true b-jet and light (*left*) and charm (*right*) flavour mis-tag rate for the main CMS b-tagging algorithms [65].

3.6 The Inclusive Vertex Finder, a b-hadron identification tool

The most widely used b-tagging techniques listed in the previous section are designed to search for a heavy flavour particle decay within a jet. As mentioned above, the hadronic showers generated by a pair of partons with small angular separation merge into a single reconstructed jet. In these cases, therefore, the two partons cannot be disentangled and their flavour cannot be determined, as shown in Fig. 3.14. Fig. 3.15 shows the transverse angular separation $\Delta\phi$ for a sample of QCD events generated by Pythia, for all true b hadrons, and for events where each of the two b hadrons gives a reconstructed jet. A significant inefficiency is observed in the region of small angular separation, proving that the collinear b-quark production, yet interesting from the physics point of view, is hence inaccessible via the standard jet-based methods. The sensitivity to the entire angular spectrum can be recovered only by using an identification tool not relying on jets, thus not suffering from their geometrical limitations: the *inclusive vertex finder* (IVF). A detailed description of the algorithm and its performance is given in this section.

3.6.1 The inclusive vertex finding algorithm

The IVF identifies b hadrons by reconstructing the secondary vertices (SV) from their decay, based on the same techniques used for the PV. It consists of a sequence of several steps, described in detail in the following subsections.

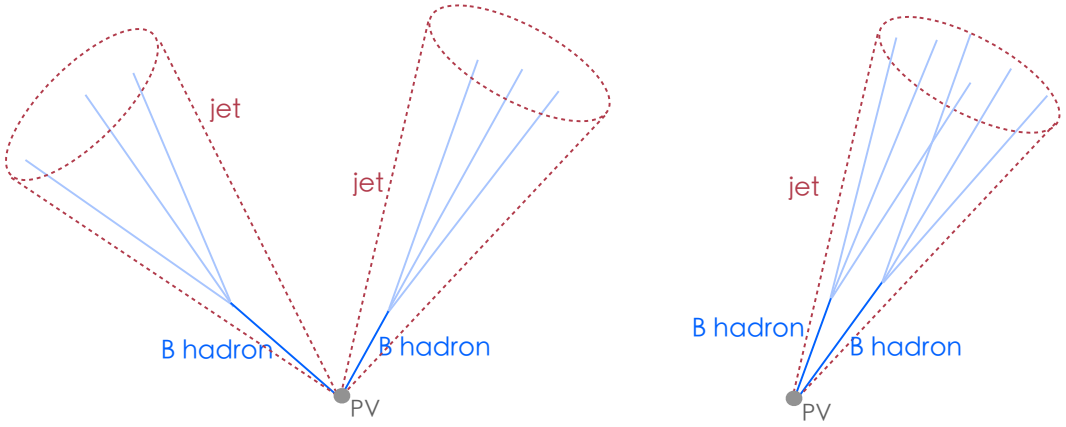

Figure 3.14: Identification of a b-hadron pair with b-tagging techniques using jets: two particles separated by a small opening angle are merged into a single jet and cannot be identified as separated objects.

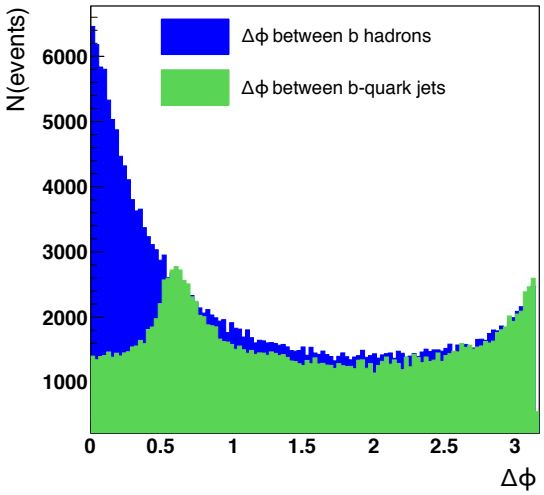

Figure 3.15: Transverse separation $\Delta\phi$ between two simulated b hadrons, and between the two corresponding reconstructed jets. A significant inefficiency is observed in the collinear region of the angular spectrum.

Nomenclature and conventions

The purpose of this section is to collect some definitions that will be extensively used in the following discussion.

The momentum and energy of a vertex are obtained as sum of the momenta and energies of

the single tracks associated to it, under the assumption that the charged tracks are pions. The vertex invariant mass is calculated from the track set as well, with the same mass assumption.

In case of multiple primary interactions, the IVF uses the first candidate in the reconstructed primary vertex collection, the one with the highest sum of the track p_T^2. Angles are computed with respect to the vertex flight direction, i.e. the vector connecting the selected PV and the SV, which well approximates the b hadron propagation direction. For particles with short lifetime, such as the Z boson, the momentum is used instead.

The particle-level object of reference in the context of this analysis is the *b hadron*. The b-hadron category includes several types of particles, such as $B^{\pm}$, B^0, B_s^0 mesons and barions containing b quarks.

The IVF performance studies and the determination of the efficiency and purity correction factors, described in Section 4.3.4 and 4.3.5 , require to define a criterion for the matching between reconstructed B candidates and simulated b hadrons. The association is performed according to a geometrical condition: the two objects are considered matched if the three-dimensional separation between their flight directions is smaller than 0.1.

Seeding and track clustering

The first stage in the IVF sequence is the identification of highly displaced tracks, providing the SV reconstruction with seeds. Only High Purity tracks with a total number of hits $N_{hits} > 8$ and $p_T > 0.8$ GeV are used, and those with a three-dimensional impact parameter value $IP > 50$ µm and a significance $s_{IP} = \sigma_{IP}/IP > 1.5$ are selected as seeds.

For each seed, the algorithm subsequently searches for additional tracks to be clustered, according to the compatibility with the hypothesis of a common origin. The compatibility is evaluated based on a set of variables describing the linear and angular separations between the seed and the track under examination. Although the cut values of these variables are configurable, the performance of the algorithm is substantially independent of any choice in this context [66]. As clusters with more than 30 associated tracks most likely correspond to the PV, they are excluded from any further processing.

Vertex fitting, merging and track arbitration

Each set of tracks selected in the clustering step is fitted to a common vertex using the adaptive vertex finder algorithm described in Section 3.4. The AVF free parameters are set to values appropriate for the reconstruction of secondary vertices [66].

As a single track can be both identified as seed and clustered to another seed, the vertices can share a subset of their tracks. To avoid duplicates, a cleaning procedure is applied, that merges pairs of vertices separated by a distance with significance smaller than 2, or having

more than 70% of tracks in common. The SV with the highest track multiplicity is kept. After this preselection, the algorithm further selects all secondary tracks with $p_T > 0.4$ GeV, with at least one hit in the pixel system, and measurements in at least four tracker layers, and evaluates their compatibility with the primary vertex, and with the selected SV. The estimation is based on the three-dimensional impact parameter with respect to the SV, to the PV, and the respective significances. Once that all the tracks are sorted and associated to the proper vertex, all the secondary vertices are refitted. For data analysis, the merging procedure is applied again with tighter requirements [66].

3.6.2 B candidate reconstruction and selection

The IVF is primarily designed to identify b hadrons. Nevertheless, it is capable of reconstructing the decay of any type of particle, provided it has a sufficiently long lifetime. This is the case of charmed hadrons, decaying mostly to kaons with a length comparable to b hadrons. On one hand, the capability of reconstructing charmed hadrons as well deteriorates the tagging performance, as it introduces a contamination from charm backgrounds. On the other hand, it significantly increases the efficiency of identifying sequential b $\rightarrow$ c $\rightarrow$ X decays, the most common. In these cases there is a high probability that at least one of the two long-lived particles gives a high quality SV passing all selection cuts.

In some cases, both the b and the charmed hadron decays are identified by the IVF, resulting in two vertices at small angular separation which can be mistaken for a quasi-collinear b-hadron pair. The B candidate producer can distinguish between SV pairs from collinear b-hadron production, and b $\rightarrow$ c $\rightarrow$ X decay chains. The algorithm sorts all reconstructed vertices according to their invariant mass, and examines them in pairs. Vertex pairs are merged into a single *B candidate* if the following conditions are fulfilled:

- The three-dimensional angular separation between the two vertices is smaller than 0.4;
- The total invariant mass is lower than 5.5 GeV;
- The angle θ between the vector connecting the PV to the closest SV and the momentum of the vertex with longer decay length is such that $\cos\theta < 0.99$, as shown in Fig. 3.16.

In the merging procedure, the lightest vertex is removed, and its tracks are associated to the other, the momentum of which becomes equal to the sum of the momenta of the two initial vertices.

At the end of the IVF sequence, the output B candidate collection is still affected by a sizeable contamination from charmed hadron decays. Secondary vertices are therefore further selected to reject the charm background and enhance the b fraction. The filter requires a vertex invariant mass $m_{SV} > 1.4$ GeV and a three-dimensional flight distance significance larger than 5.

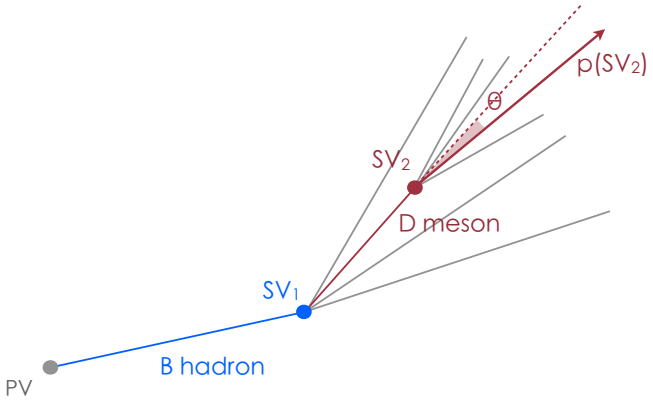

Figure 3.16: Skematic illustration of the identification of SV pairs from sequential b $\rightarrow$ c $\rightarrow$ X decays.

For the analysis described in this thesis, B candidates are rejected if $p_T < 8$ GeV and $|\eta| > 2$. The purpose of the η cut is to select only the central part of the CMS tracking region, which guarantees better track and vertex reconstruction performance.

3.6.3 b-hadron identification performance

Most of the selection cuts applied in the IVF sequence and listed in the previous sections are configurable. A detailed cut optimisation was performed during the development of the tool, and is documented elsewhere [66]. The performance of the IVF as b-hadron tagger is assessed in a detailed study based both on Monte Carlo simulation and on data. The results are collected in this section and in Section 4.3.4.

Angular resolution and momentum reconstruction

The results shown in this section are obtained using the simulated MC sample of Drell-Yan events with additional jets described in Section 4.2, filtered by requiring the presence of at least one b-quark. As some dependence of the performance is expected from the event dynamics, they might not be directly applicable to a more general case.

For the purpose of an angular correlation measurement, an important parameter is the resolution of the single b-hadron flight direction, or equivalently of the angle between b hadrons. Fig. 3.17 shows the the ΔR and $\Delta\phi$ separation (defined in Section 2.2) between two reconstructed B candidates (ΔR_{BB}(reco) and $\Delta\phi_{BB}$(reco)), and between two simulated b hadrons (ΔR_{BB}(sim) and $\Delta\phi_{BB}$(sim)), in events with exactly two b-hadron candidates and two true b hadrons. The bin size is the same as used in the analysis presented in Chapter 4. About 99% of the events are clustered along the diagonal, showing no bias of the measured an-

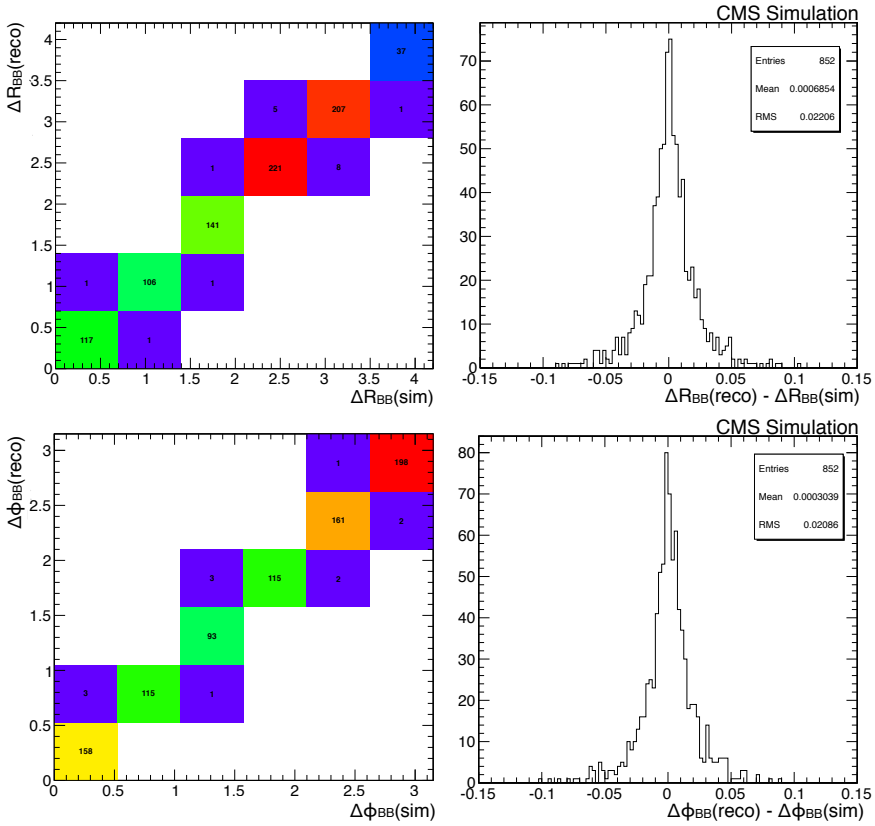

Figure 3.17: Angular resolution of the IVF method: correlation between reconstructed and true ΔR (*top left*), and $\Delta\phi$ (*bottom left*), and projection onto the diagonal (*top* and *bottom right*). No bias is observed in the measured angular separations with respect to the simulated values. The transverse and three-dimensional resolution, estimated as RMS of the distribution, is approximately 0.02.

gles with respect to the true values. The projection onto the diagonal, i.e. the difference $\Delta R_{BB}(\text{sim}) - \Delta R_{BB}(\text{reco})$, yields a distribution that is in first approximation Gaussian. The width gives an estimate of the angular resolution, which is observed to be about 0.02 for both $\Delta\phi$ and ΔR. This value has to be compared to the bin width in an angular differential cross section measurement. In the analysis described in this thesis, the bin width is constrained by the limited size of the data sample, and is between 0.6 and 0.8 in ΔR, and between 0.5 and 0.6 in $\Delta\phi$. The resulting migration of events from one bin to the adjacent ones is lower than 4%, which allows to avoid a full unfolding.

The measurement of energy and momentum of an IVF B candidate is rather poor. As only charged particle tracks are included in the vertex reconstruction, and consequently in the computation of the total momentum, the fraction carried by neutral particles is lost. On average,

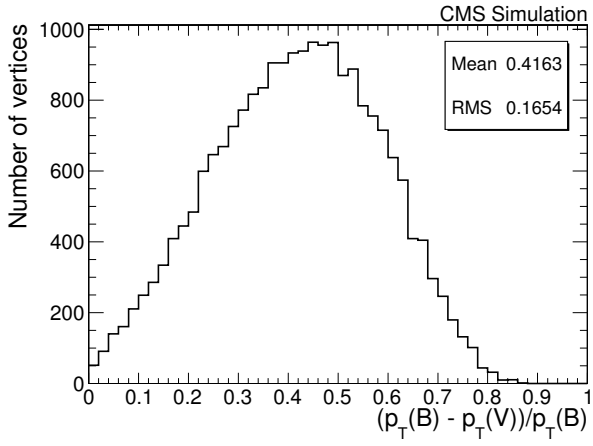

Figure 3.18: Transverse momentum resolution of B candidates reconstructed with the IVF, evaluated as difference between the p_T of the true simulated b hadron and of the corresponding reconstructed B candidate, divided by the simulated b hadron p_T. The plot is for events with at least one B candidate.

this corresponds to approximately 60% of the true b hadron momentum, with a large spread, as shown in Fig. 3.18.

b-hadron identification efficiency and data/MC scale factors

A differential cross section measurement requires to evaluate the efficiency for identifying a b-hadron pair as a function of all the variables of interest. Such a study is documented in Section 4.3.4, as the results are strongly dependent on the event kinematics, and are therefore specific of the chosen event topology. This section describes a measurement of the IVF absolute b hadron identification efficiency, and the comparison with the Monte Carlo simulation.

The efficiency can be evaluated on data with the so-called muon p_T^{rel} method, based on semi-leptonic decays of b hadrons with muons in the final state, and extensively used for the calibration of standard b-tagging algorithms [65]. Although the IVF is independent from jet reconstruction, the use of jets for the efficiency measurement with data is unavoidable. The method exploits the fact that, due to the relatively heavy mass of b hadrons, leptons from their decay have on average a larger transverse momentum with respect to the jet axis (p_T^{rel}), than in charm and light decays. A sample of events with exactly two jets is selected: one jet — the so-called *muon jet* — is required to contain a muon, while the other one is b-tagged by the TC algorithm in the working point that provides the highest b purity. The secondary vertices reconstructed by the IVF are matched to muon jets if their flight direction is within a $\Delta R < 0.2$ cone around the jet axis. This procedure is valid under the reasonable assumption that the difference between vertices within and outside a jet is correctly modelled by the simulation.

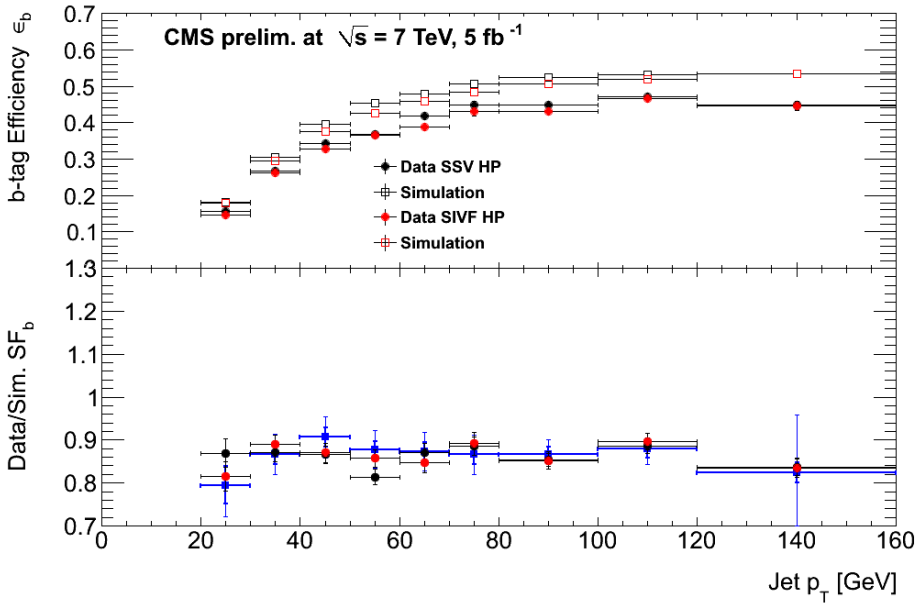

Figure 3.19: b-hadron identification efficiency measured with data with the muon p_T^{rel} method, as a function of the jet p_T. The observed values are compared to the prediction by the simulation to extract the data/MC scale factors needed for physics analysis.

Two samples are created, with jets passing (*tagged*) or failing (*untagged*) the b-hadron identification requirements. The fraction of b, charm and light flavour jets is extracted from a binned maximum likelihood fit of the p_T^{rel} distribution. The efficiency is calculated from the fit results using the following formula:

$$\epsilon_b^{tag} = \frac{f_b^{tag} \cdot N_{data}^{tag}}{f_b^{tag} \cdot N_{data}^{tag} + f_b^{untag} \cdot N_{data}^{untag}} \tag{3.4}$$

where f_b^{tag} (f_b^{untag}) is the fitted fraction of b-jets in the tagged (untagged) jet sample, and N_{data}^{tag} (N_{data}^{untag}) the absolute number of tagged (untagged) jets. As some dependence is expected on the event kinematics, the measurement is performed in bins of the muon jet p_T, as shown in Fig. 3.19. The IVF efficiency curve obtained from data is compared to the Monte Carlo prediction, as well as to the results for the SSV b-tagging algorithm. The observed dependence on p_T is well reproduced by the simulation, while there is an overall disagreement of approximately 7% between data and the simulation, visible in the bottom panel of Fig. 3.19. In the context of cross-section measurements, where correction factors are needed to extract the hadron-level quantities, the IVF efficiency values obtained from the simulation have to be rescaled to account for this discrepancy, before being used to correct the data.

IVF dependence on pileup

To be applied on data collected at large instantaneous luminosities, the reconstruction tools have to guarantee a certain robustness against pileup. The independence of the IVF performance, in particular of the efficiency, on the number of proton-proton interactions per bunch crossing is controlled using data and MC simulation. Fig. 3.20 shows the ratio between the number of events with exactly one reconstructed B candidate and with no B candidate, in bins of the primary vertex multiplicity. This variable is indicative of the single b-hadron tagging efficiency, and allows for differential studies when the statistics is not sufficient for a data-driven analysis similar to the one described in Section 3.6.3. Two conclusions can be reached from

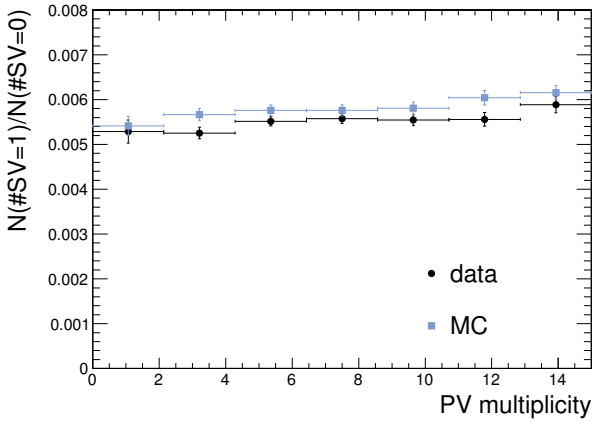

Figure 3.20: Ratio between the number of events with exactly one reconstructed B candidate, and with no B candidates, as function of the primary vertex multiplicity. Data is compared to the Monte Carlo prediction.

Fig. 3.20:

- The comparison between data and the simulation shows a discrepancy of about 10%, with data being lower than the prediction. This observation is in agreement with the measurement of the data/MC efficiency scale factors based on the p_T^{rel} method, and does not depend on the pileup;

- The ratio shows no significant trend as a function of the PV multiplicity. The algorithm is sufficiently robust against pileup to be used for physics analysis.

3.7 Summary and personal contributions

This chapter gave an overview of the event reconstruction software at CMS. A particular attention was addressed to the aspect of relevance for the analysis described in this thesis: hit detection in the pixel and silicon strip systems, tracking and vertex reconstruction, and heavy flavour particle identification.

My contribution consists of the pixel hit resolution measurements with the overlap method, described in Section 3.1, and of a study of the performance of the inclusive vertex finder, based on simulation, in topologies with a Z boson and two b hadrons. This analysis was aimed at commissioning the IVF for 2011 high luminosity data.

CHAPTER

FOUR

STUDY OF THE PROCESS pp $\rightarrow$ Zb$\bar{\text{b}}$X

This chapter presents the measurement of the angular correlations and the total cross section of the process pp $\rightarrow$ Zb$\bar{\text{b}}$X, based on a data sample corresponding to an integrated luminosity of 5.15 fb^{-1} at a center-of-mass energy $\sqrt{s} = 7$ TeV. Four variables are used to characterise the angular correlations:

- The three-dimensional angular separation between the b-hadron flight directions (ΔR_{BB}) is the reference variable of this analysis, as it is sensitive to the Zb$\bar{\text{b}}$ production mode. It allows to probe the modelling of the qq $\rightarrow$ Zb$\bar{\text{b}}$X and qg $\rightarrow$ Zb$\bar{\text{b}}$X subprocesses, with Feynman diagrams containing the g $\rightarrow$ b$\bar{\text{b}}$ splitting vertex, disentangling them from the subprocess gg $\rightarrow$ Zb$\bar{\text{b}}$X. The two categories populate different regions of the ΔR_{BB} spectrum, with the gluon splitting being predominantly responsible for the production of b-hadron pairs at small angular separation ($\Delta R_{BB} < 1$);

- The angular separation between the b-hadron flight directions in the transverse plane ($\Delta\phi_{BB}$) complements the information provided by ΔR_{BB}. The subprocesses characterised by a gluon splitting vertex are dominant in the region $\Delta\phi_{BB} < 0.75$;

- The three-dimensional angular separation between the Z momentum and the flight direction of the closest b hadron (minΔR_{ZB}). As ΔR_{BB} and $\Delta\phi_{BB}$, it constitutes a test for the modelling of the subprocesses including the g $\rightarrow$ b$\bar{\text{b}}$ splitting, which are dominant in the big separation region (min$\Delta R_{ZB} > 3.2$). Furthermore, minΔR_{ZB} is sensitive to the emission of additional QCD radiation in the final state, modifying the energy balance between the Z boson and the b hadrons;

- The asymmetry between the Z momentum and the b-hadron flight directions (A_{ZBB}) is defined as

$$A_{ZBB} = \frac{\max\Delta R_{ZB} - \min\Delta R_{ZB}}{\max\Delta R_{ZB} + \min\Delta R_{ZB}}, \tag{4.1}$$

where maxΔR_{ZB} is the three-dimensional angular separation between the Z momentum and the flight direction of the farthest b hadron. A_{ZBB} is approximately zero when the b hadrons are emitted symmetrically with respect to the Z boson momentum, and assumes nonzero values in the presence of additional QCD radiation emissions. Together with minΔR_{ZB} it allows to test the validity of the calculations at higher order in the perturbative series.

The differential cross sections are evaluated with no requirement on the Z boson p_T (p_T^Z), and for $p_T^Z > 50$ GeV, to explore the evolution of the event dynamics in boosted Z configurations. The requirement on the Z boson p_T^Z is expected to enhance the relative contribution from subprocesses with the gluon splitting g $\rightarrow$ b$\bar{\text{b}}$ vertex. Fig. 4.1 shows the differential Zb$\bar{\text{b}}$ production cross section as a function of the four angular variables, for the two p_T^Z regions considered, as predicted by the four-flavour MadGraph generator presented in Section 1.3.5. A requirement applied to the initial state particle type allows to isolate the contribution from subprocesses with a gluon splitting vertex g $\rightarrow$ b$\bar{\text{b}}$, and to show the phase space regions in which it is dominant.

The total Zb$\bar{\text{b}}$ production cross section is also measured, in four kinematical regions: $p_T^Z > 0, 40, 80$ and 120 GeV.

All cross sections are measured in the phase space defined by the following requirements:

- Lepton $p_T^\ell > 20$ GeV and $|\eta^\ell| < 2.4$;
- Dilepton invariant mass $81 < M_{\ell\ell} < 101$ GeV;
- b-hadron $p_T^B > 15$ GeV and $|\eta^B| < 2.0$.

The lepton momentum is corrected for final state radiation, i.e. the momentum of all the generator-level photons in a $\Delta R < 0.1$ cone around the lepton propagation direction is added to the lepton momentum. The b-hadron category includes any long-lived hadron containing one valence b-quark, and in particular the B^+, B^0, B_s, B_s^0, B_c^+ mesons and their antiparticles, and the Λ_b^0 baryon. In general b hadrons have masses between 5.28 and 6.28 MeV, and a lifetime of the order of 1.5×10^{-12} s, resulting in decay lengths of about 500 μm. Their long lifetime provides a handle for the b hadron identification. The decay into hadrons with a c quark is highly dominant.

The first analysis step consists of the identification of events containing a Z boson decay into a muon or electron pair. The selection is performed at the online level, by the dilepton trigger filters at L1 and HLT, and subsequently to the offline event reconstruction. Standard criteria, extensively used within the CMS experiment, are used for the lepton identification. Pairs of b hadrons are reconstructed through their displaced decay vertices, using the Inclusive Vertex Finder technique discussed in Section 3.6. Details of the event selection are reported in Section 4.1. Section 4.2 gives an overview of the event dynamics, for the topology of interest,

and after the main selection steps. The relevant lepton and b-hadron kinematic distributions measured at the detector level are compared to the Monte Carlo prediction, for a validation of the simulation. The measurement of the total and differential cross sections is discussed in Section 4.3. Section 4.4 is dedicated to the description of the theoretical predictions to which the data is compared to, and their uncertainties. The results are presented in Section 4.5.

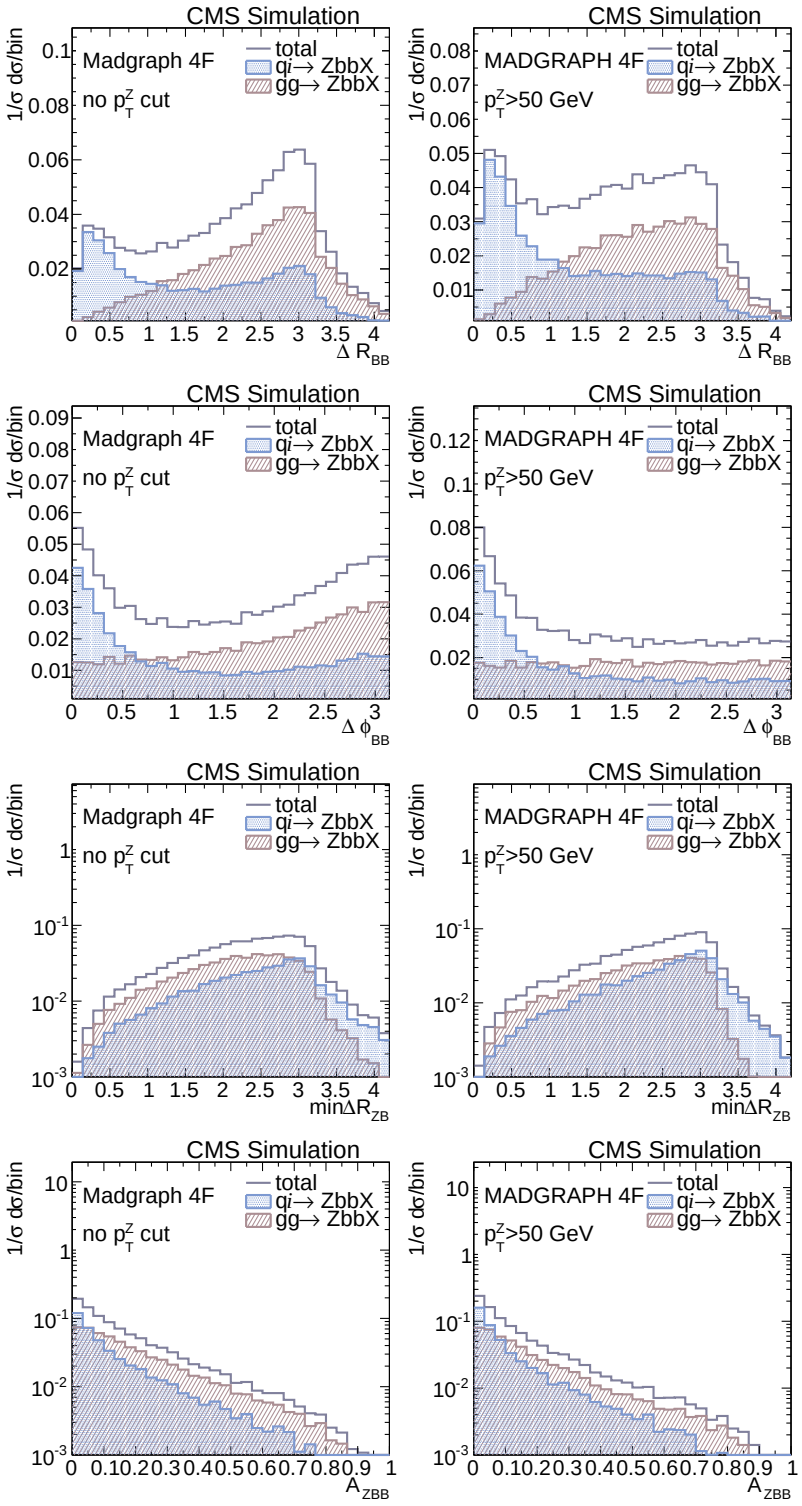

Figure 4.1: Differential Zbb̄ production cross section as a function of the four angular variables described in the text, ΔR_{BB} (*first row*), $\Delta\phi_{BB}$ (*second row*), $\min\Delta R_{ZB}$ (*third row*), and A_{ZBB} (*fourth row*), with no cut on p_T^Z (*left*) and for $p_T^Z > 50$ GeV (*right*), by MadGraph.

4.1 Event selection

4.1.1 Trigger selection

The first step of the event selection is applied at the online trigger level. The analysis uses dilepton HLT trigger paths, seeded by L1 muon and ECAL triggers. In 2011, HLT requirements have been periodically updated to keep the trigger rates under control despite the rapidly increasing instantaneous luminosity. Higher momentum thresholds and tighter isolation cuts were applied at each update. Trigger paths with looser requirements and higher rates were prescaled (only a fraction of randomly chosen events passing the filters were retained and recorded) reducing the efficiency by the same prescale factor. The unprescaled paths with the loosest cuts for each run period are therefore chosen for this analysis.

Three different dimuon trigger paths are used, all requiring the presence of two muons with high transverse momentum. The p_T thresholds were raised from 7 GeV for both legs, to 13 and 8 GeV, to 17 and 8 GeV, with the increasing luminosity. The dielectron trigger paths impose the presence of two electron candidates, with p_T higher than 17 and 8 GeV. Loose cuts are imposed on the isolation — i.e. the energy and momentum from the hadronic activity surrounding the leptons— and on the electron identification variables described in Section 4.1.2.

4.1.2 Offline lepton identification and selection

Muons

Muons are identified using the particle flow (PF) algorithm [64]. A muon candidate is kept if the following conditions are fulfilled:

- It is both a global and a tracker muon, according to the definitions given in Section 3.5. The track from the inner tracker is matched to at least two segments in the muon stations;
- The transverse momentum p_T^μ is larger than 20 GeV. The offline momentum cut has to be sufficiently tighter than the threshold applied at the HLT level, to avoid the region of steep increase in the trigger efficiency turn-on curve, and to ensure the full efficiency of the online filter;
- The pseudorapidity is in the range $|\eta^\mu| < 2.4$;
- The transverse impact parameter d_0^μ of the muon trajectory with respect to the primary vertex satisfies the condition $d_0^\mu < 200$ µm. This requirement selects muons originated in the primary collision only, and helps in rejecting leptons from decays of long-lived hadrons;

- At least 8 hits are detected in the pixel and silicon strip detectors, and the χ^2 of the tracker track fit is lower than 10;
- The offline PF muon is linked to one of the HLT muons that caused the trigger to fire. The two objects are matched if their momentum vectors satisfy the condition $\Delta R < 0.3$;
- It is sufficiently isolated, i.e. the total momentum and energy surrounding the lepton is lower than a given threshold. This requirement rejects muons from semi-leptonic decays of hadrons, mostly pions and kaons, which are usually embedded in a jet. It also reduces the chance that a muon from the decay of a B hadron is used in the construction of the Z boson candidate. An *isolation* variable is built as scalar sum of the contributions from PF photons, charged and neutral hadrons, divided by the lepton transverse momentum, according to Eq. 4.2

$$I = (\sum^{charged\ had} p_T + \sum^{neutral\ had} p_T + \sum^{photon} E_T)/p_T^{\mu}, \tag{4.2}$$

  where the sum is extended to all objects within a cone centered around the lepton momentum direction. Muons are selected if $I < 0.2$ in a $\Delta R < 0.4$ cone. According to the definition above, the isolation variable is extremely sensitive to pileup. Corrections are therefore applied, in order to mitigate the dependence of the efficiency of the isolation cut on the LHC instantaneous luminosity. The idea is to estimate the energy and momentum of particles associated to the selected primary interaction only, and to reject the rest. For the charged hadrons a full geometrical matching can be performed using the track information. The neutral hadron and photon terms ($\sum^{neutral\ had.} p_T$ and $\sum^{photon} E_T$ in Eq. 4.2) are corrected by subtracting the estimated fraction of energy and momentum originating from pileup interactions. This is evaluated as half of the charged hadron energy deposits in the calorimeter.

Electrons

This analysis uses the GSF electrons defined in Section 3.5, while particle flow is applied in the calculation of the isolation corrections. Good electron candidates are selected by a standard identification sequence, which includes cuts on several variables: the energy-momentum and spatial match between the tracker track and the ECAL supercluster, the supercluster width, and the hadronic leakage, defined as the ratio between the energy deposit in ECAL and HCAL. All cuts are configurable. Several working points are commissioned and available for data analysis, identified by a number that specifies the corresponding efficiency. This work uses the working point giving a 85% efficiency for prompt electrons. More details about electron identification are given in [62].

Additional requirements are applied to select electrons from Z boson decays. An electron candidate is kept if:

- The transverse momentum p_T^e is above 20 GeV. This value is sufficiently larger than the threshold applied at the HLT level to ensure full trigger efficiency;
- The pseudorapidity is in the range $|\eta^e| < 2.4$. The interval $1.442 < |\eta^e| < 1.566$, corresponding to the low-performance region in the transition between ECAL barrel and endcap, is excluded;
- The transverse impact parameter d_0^e satisfies the condition $d_0^e < 200$ µm;
- The offline electron can be linked to one of the HLT electrons. The matching criterion is the same as for muons;
- It is isolated, where the isolation variable is defined as for muons. The sum is extended to PF objects in a $\Delta R < 0.3$ cone around the electron momentum. As in the muon case, some corrections have to be applied to ensure a sufficient robustness of the isolation cut against pileup. Only charged hadrons associated to the selected primary interaction are included in the sum. For the neutral component, the average energy density ρ from pileup interactions within the isolation cone is estimated as the median of the jet energy distribution divided by the effective cone area. The density ρ is subtracted from the $\sum^{neutral\ had.} p_T + \sum^{photon} E_T$ term of Eq. 4.2.

Z candidate

Only events with exactly two opposite-sign leptons of the same flavour passing all selection cuts are retained. Events with three selected leptons are rejected. The invariant mass of the dilepton pair is required to be in the narrow window centred in the Z boson mass ($81 < M_{\ell\ell} < 101$ GeV) to suppress the contamination from background processes without a Z in the final state ($t\bar{t}$ production).

4.1.3 B-candidate selection

The sample of events passing the dilepton requirements is expected to be heavily dominated by Drell-Yan production. Given the small cross section of the associated production of Z and heavy flavor particles, only a very small fraction of events — less than 1% — contain B hadrons. In order to enhance this component, and to suppress the contamination from Z produced in association with light and charm quarks, the inclusive vertex finder is applied, in the configuration described in Section 3.6. A B candidate is retained if:

- The transverse momentum p_T^B is larger than 8 GeV. Assuming the visible fraction of the momentum to be approximately 40%, this value approximately corresponds to $p_T^B > 15$ GeV in the phase space acceptance definition;

- The pseudorapidity η^B is $|\eta^B| < 2$. This cut selects secondary vertices and tracks detected in the central part of the tracker system, which guarantees the best performance;
- The secondary vertex mass m^B is heavier than 1.4 GeV;
- The SV three-dimensional flight distance significance s^B is larger than 5;
- At least 3 tracks are associated to the SV.

The requirements on m^B, s^B and track multiplicity reduce the contamination from mis-tagged charmed mesons. Only events with exactly two reconstructed B candidates satisfying the selection criteria listed above are retained for further analysis.

4.2 Event properties and comparison with Monte Carlo simulation

The total number of data events passing the selection requirements listed in the previous section is 553, 330 in the dimuon, and 223 in the dielectron channel. A detailed survey of the event dynamics is performed, not only for the sample selected for the cross section measurement, but also for some of the intermediate selection stages. At every step, the total event yields and the distributions of the relevant kinematic variables observed in data are compared to the fully simulated and reconstructed MC, to validate the predictions at the detector level.

In order to use a simulated Monte Carlo sample for this purpose, the stable and meta-stable particles constituting the hadronisation output have to be further propagated and left to interact with the detector. The transition between the jets of particles originated at the proton-proton interaction point and the pattern of signals collected by the different detector subsystems is handled by the GEANT4 package [67]. The program includes an accurate modelling of the particle interaction with matter and of long-lived particle decays, and uses a detailed description of the detector structure. The simulated set of signals is finally processes by the same event reconstruction software that is applied to the data, described in Chapter 3. As the simulation of the particle detection and the event reconstruction are computationally demanding, for many MC samples the events are propagated to the hadronisation stage only. Such samples are nevertheless useful for the comparison to hadron-level observables in data.

The MC expectation is built by summing the contributions from the main processes involved: Drell-Yan, diboson and $t\bar{t}$ production. Table 4.1 lists the signal and background samples used, with the corresponding size and production cross section.

In the figures referring to the inclusive Z selection, the Drell-Yan (DY) component corresponds to the MG 5F sample, described in Section 1.3.5. The inclusive sample is divided into three subsamples: the DY+light, the DY+charm and the DY+b-quark components, filtered by requiring or vetoing the presence of charm and b quarks. The plots for the final analysis selection (from Fig. 4.6 to 4.11) are obtained by replacing the Z+b-quark component of MG 5F with the MG

Final state	Generator	cross section (pb)	luminosity (pb^{-1})
$Zb\bar{b}$	MadGraph 4F	44	3417686
Drell-Yan+jets	MadGraph 5F	3048	11897
$t\bar{t}$	MadGraph	157	240196
ZZ	Pythia	4.3	994498

Table 4.1: List of Monte Carlo samples used in this work, with the corresponding cross section and equivalent integrated luminosity.

4F sample. Each process is rescaled by a global normalisation factor that takes into account the data integrated luminosity $\mathcal{L}$, the cross section σ and the MC sample size N reported in Table 4.1, according to the formula $n = \sigma \cdot \mathcal{L}/N$. In specific cases, explicitly specified in the text, the total MC yield is normalised to the number of events observed in data, allowing for a pure shape comparison. In general, three additional rescaling factors have to be applied to the simulation, on an event-by-event basis:

- The trigger filters of the chosen paths are not implemented in the simulation. The effect of the trigger selection has therefore to be introduced a posteriori, by re-weighting the events for the filter inefficiencies. The efficiency of the chosen trigger paths is measured with data, using the tag & probe method described in Section 4.3.2, in bins of the p_T and η of each of the legs [68]. The resulting values are applied event-by-event according to momentum and pseudorapidity of the two leptons;

- Small discrepancies, of the order of 1%, are observed between the dilepton reconstruction and selection efficiencies extracted from data and the values predicted by the simulation. The measurement is performed with a tag & probe technique, and is described in Section 4.3.2. In order to consistently compare the yields, these discrepancies have to be corrected for. Thus, a set of event-by-event scale factors, defined as the ratio between measured and simulated efficiency value, are applied, depending on the kinematics of the two leptons [68].

- Additional correction factors — the *pileup reweighting* coefficients — are applied to the MC histograms and yields, on an event-by-event basis. For technical reasons, mostly related to the impossibility to keep up with the rapidly changing instantaneous luminosity, all the MC samples at CMS are generated with a primary vertex multiplicity distribution that does not reproduce the data. As the performance of the event reconstruction and of the selection criteria depends on the pileup conditions, the simulated events have to be rescaled for the data/MC comparison to be consistent. Each event is re-weighted according to the PV multiplicity, the set of weights being calculated before any analysis-specific selection, by dividing the PV multiplicity histogram implemented in the simulation and the expected one. The latter is estimated from data, taking into account the instanta-

neous and integrated luminosity of all good LHC runs, and the proton-proton inelastic cross section, which determine the number of hard scattering interactions in a single bunch crossing. The effect of a possible overlap with the activity from the previous or the following bunch crossing — the *out-of-time* pileup — due to the finite response time of some subdetectors, is also accounted for. The data/MC agreement achieved in the reconstructed PV multiplicity distribution after this procedure is shown in Fig. 4.2, for the sample of events satisfying the dilepton selection requirements listed in Section 4.1.2. The residual discrepancies can be due either to the uncertainty on the proton-proton inelastic scattering cross section, or to an imperfect description of the out-of-time pileup. The effect of this disagreement on the cross-section measurement is however expected to be negligible, as discussed in Section 4.3.6.

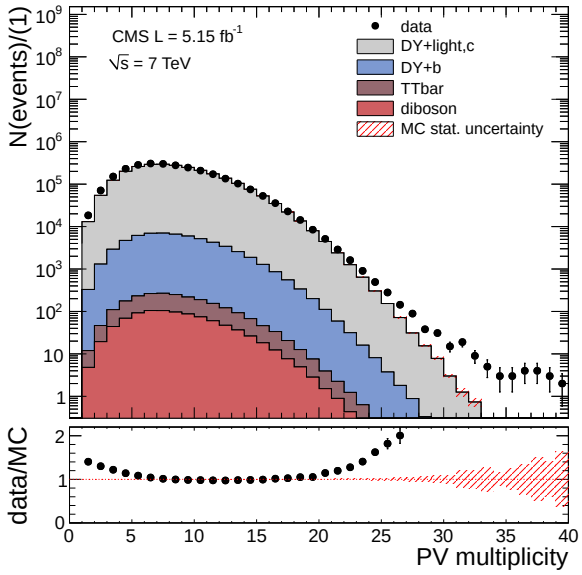

Figure 4.2: Reconstructed primary vertex multiplicity distribution after the pileup reweighting procedure described in the text.

The yields of observed and simulated events are compared in Table 4.2 and 4.3, for the dimuon and dielectron channels respectively, at three stages of the selection: after the Z candidate selection, and with the additional requirement of exactly one or two reconstructed B candidates. The table shows also the number of events observed in the upper Z sideband, $M_{\ell\ell} > 101$ GeV, which is largely dominated by $t\bar{t}$ production. Since this process is well known, and reliably reproduced in the MC simulation, the sideband is considered as a control region for this analysis. Fig. 4.3 shows the B-candidate multiplicity in events passing the dilepton selection cuts. While an excellent agreement is observed after the dilepton selection, a discrepancy between 10 and 15% is found in the number of expected and selected events with one or two reconstructed B candidates. The difference is related to in the b-hadron tagging efficiency in data

and in the simulation, evaluated as described in Section 3.6, and accounted for by the IVF scale factors applied for the hadron-level cross-section estimation.

	$81 < M_{\mu\mu} < 101$ GeV	$81 < M_{\mu\mu} < 101$ GeV 1 B	$81 < M_{\mu\mu} < 101$ GeV 2 B	$M_{\mu\mu} > 101$ GeV 2 B
data	1591140 ± 1261	7598 ± 87	330 ± 18	248 ± 16
DY+light	1068565 ± 726	138 ± 8	0.6 ± 0.6	0 ± 0
DY+c	399812 ± 443	1267 ± 25	1 ± 0.9	0.7 ± 0.6
DY+b	80013 ± 198	6108 ± 55	243 ± 11	10 ± 2
$t\bar{t}$	342 ± 3	344 ± 3	89 ± 2	250 ± 2
diboson	692 ± 2	47 ± 1	6.9 ± 0.2	0.4 ± 0.1
sum MC	1549423 ± 873	7904 ± 61	341 ± 11	260 ± 3
data/MC	1.03 ± 0.00	0.96 ± 0.01	0.97 ± 0.06	0.95 ± 0.06

Table 4.2: Measured and simulated event yields in the dimuon channel. The MC events are rescaled according to a weight that accounts for the trigger efficiency, for the data/MC dilepton selection efficiency scale factors, and for the pileup. The IVF efficiency scale factor is not applied at this stage.

	$81 < M_{ee} < 101$ GeV	$81 < M_{ee} < 101$ GeV 1 B	$81 < M_{ee} < 101$ GeV 2 B	$M_{ee} > 101$ GeV 2 B
data	1110343 ± 1054	5655 ± 75	223 ± 15	219 ± 15
DY+light	779361 ± 657	94 ± 7	0 ± 0	0 ± 0
DY+c	295573 ± 404	917 ± 22	6.4 ± 1.8	0 ± 0
DY+b	59870 ± 182	4763 ± 51	198 ± 10	9.8 ± 2.3
$t\bar{t}$	280 ± 3	285 ± 3	76 ± 1	208 ± 2
diboson	371 ± 2	26 ± 1	2.9 ± 0.2	0.20 ± 0.0
sum MC	1135455 ± 793	6085 ± 56	284 ± 11	218 ± 3
data/MC	0.98 ± 0.00	0.93 ± 0.02	0.78 ± 0.06	1.00 ± 0.07

Table 4.3: Measured and simulated event yields in the dielectron channel. The MC events are rescaled according to a weight that accounts for the trigger efficiency, for the data/MC dilepton selection efficiency scale factors, and for the pileup. The IVF efficiency scale factor is not applied at this stage.

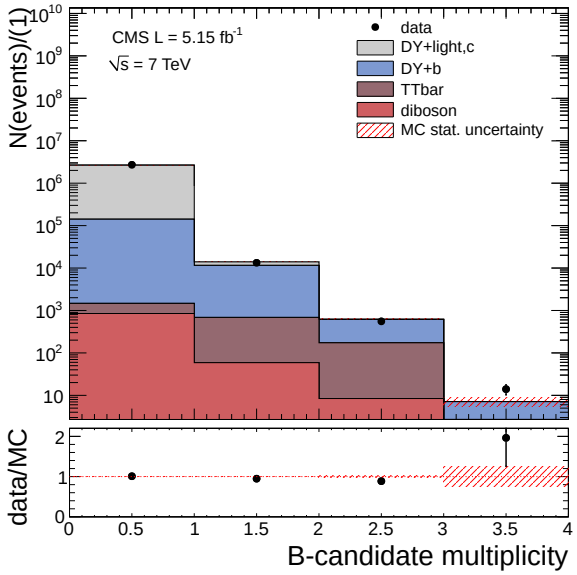

Figure 4.3: B-candidate multiplicity distribution for events passing the dilepton selection. The MC events are rescaled according to a weight that accounts for the trigger efficiency, for the data/MC dilepton selection efficiency scale factors, and for the pileup. The IVF efficiency scale factor is not applied at this stage.

4.2.1 Inclusive Drell-Yan sample

An inclusive sample of Drell-Yan events is selected by applying the dilepton cuts listed in Section 4.1. At this selection stage, the contamination from diboson and $t\bar{t}$ production is negligible. Due to its smaller cross section, the associated $Zb\bar{b}$ production constitutes a very small fraction of the total Drell-Yan yield, below 1%.

The leading and subleading lepton p_T distributions are visible in Fig. 4.4. Fig. 4.5 shows also the dilepton invariant mass distribution, for the dimuon and dielectron final states. While the observed yield is in agreement with the MC prediction, the Z invariant mass shape seems not to be correctly reproduced by the simulation. The effect is present in both dilepton channels. Although several hypotheses have been considered — either related to a bad description of the calorimeter response, or to a mismodelling of the initial and final state radiation — none of them is capable of fully accounting for the observed effect. Fig. 4.5 shows the Z boson candidate p_T as well, for the dimuon and dielectron channels. The agreement between data and prediction is excellent over the entire spectrum.

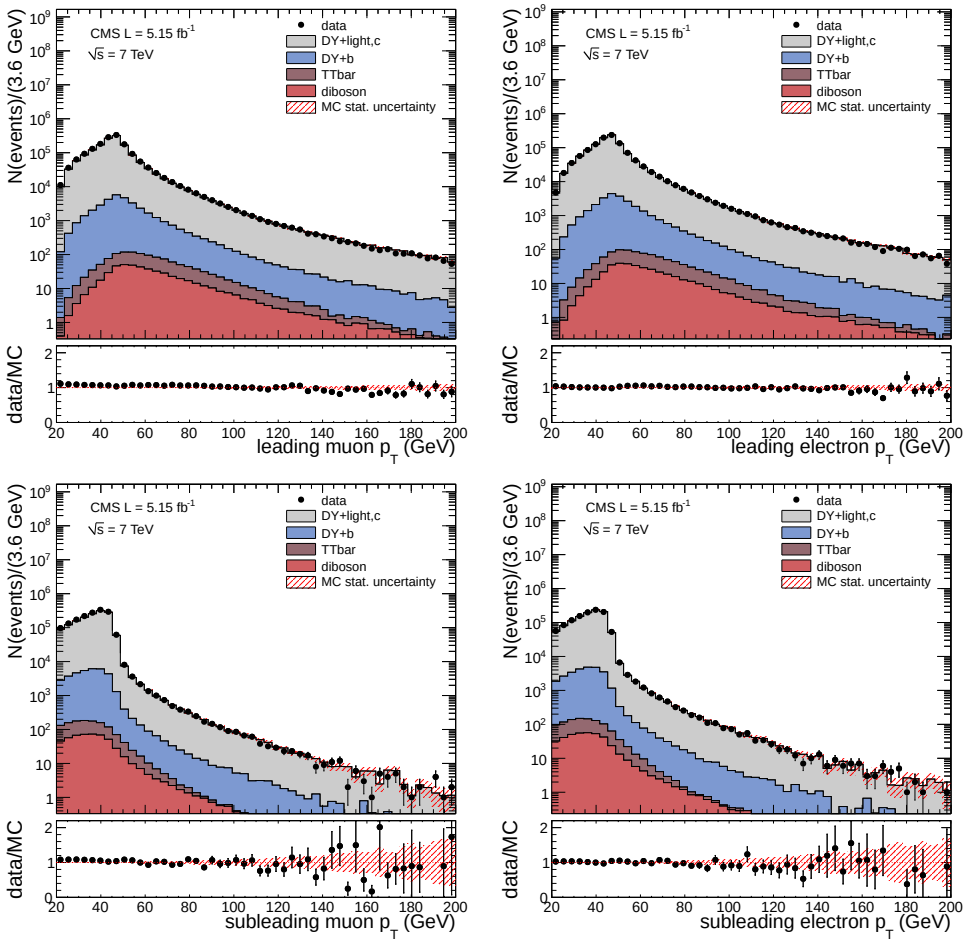

Figure 4.4: Transverse momentum for the leading (*top*) and sub-leading (*bottom*) lepton, for the dimuon (*left*) and dielectron (*right*) final states. Data (full circles) are compared to the MC prediction (stacked histograms), where each process is rescaled by the corresponding cross section.

4.2.2 Drell-Yan + two B candidate sample

The capability of the MC simulation to correctly describe the dynamics of the $Zb\bar{b}$ sample, as well as the specific properties of the secondary vertices, is of great relevance. Although most of the correction factors needed for the computation of the hadron-level cross sections are extracted from data, or at least validated using data-driven methods, the estimation of the IVF-related correction terms as a function of the angular variables relies on the simulation. Since these factors are heavily dependent on the b-hadron kinematics and on the SV properties, it is crucial to validate the MC simulation with the data.

As seen in Table 4.2 and 4.3, a disagreement of approximately 15% in observed in the expected

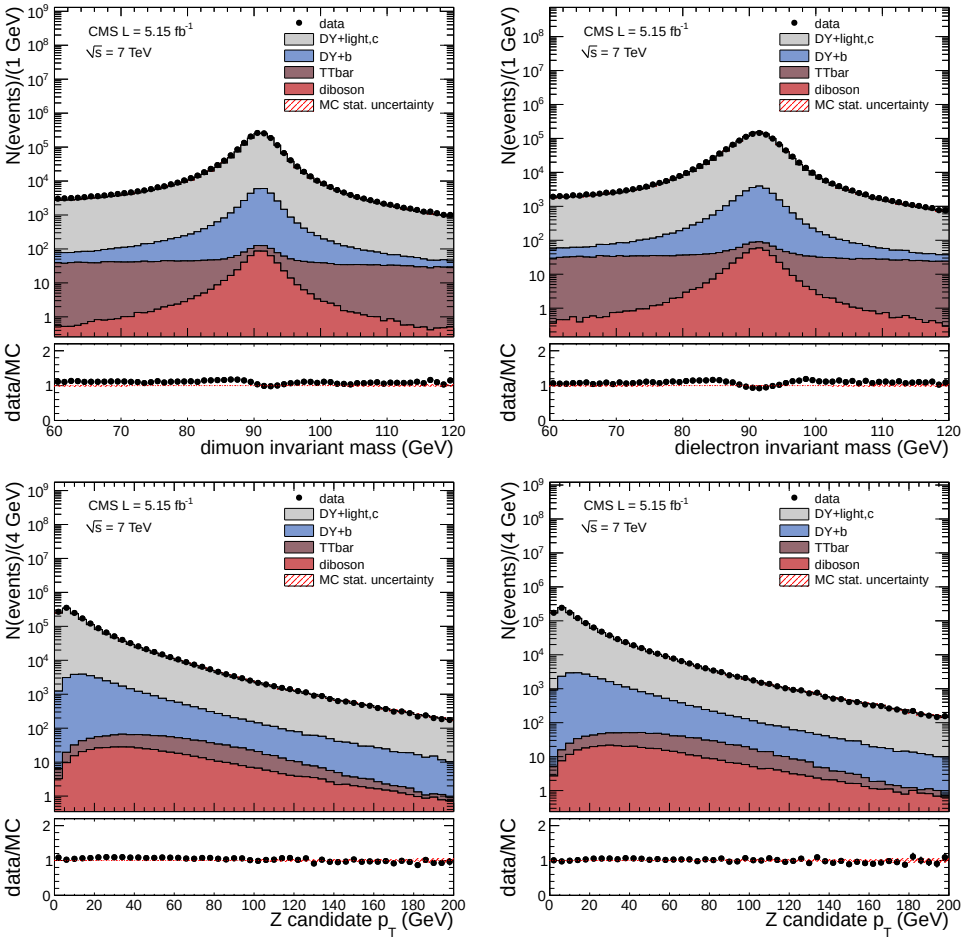

Figure 4.5: Invariant mass of the dilepton pair (*top*), and transverse momentum of the Z boson candidate (*bottom*), for the dimuon (*left*) and dielectron (*right*) final states. The invariant mass is shown in the wider range $60 < M_{\ell\ell} < 120$ GeV. Data (full circles) are compared to the MC prediction (stacked histograms), where each process is rescaled by the corresponding cross section.

and simulated yields, caused by the difference in the IVF b-hadron identification efficiency in data and in the simulation. In order to facilitate a comparison of the shapes, without any bias from the normalisation discrepancy, in the figures of this section the simulation is rescaled to reproduce the data yield. The event-by-event rescaling accounting for the dilepton trigger and offline selection efficiency is applied as described in the general introduction. The fraction of Zb$\bar{\text{b}}$ and t$\bar{\text{t}}$ is rescaled to the result of a fit of the dilepton invariant mass distribution, described in Section 4.3.1.

Fig. 4.6 shows the dilepton invariant mass, for the dimuon (*left*) and for the dielectron (*right*)

final states. A wider mass range is chosen to illustrate the effectiveness of the $81 < M_{\ell\ell} < 101$ GeV requirement for suppressing the $t\bar{t}$ background with a negligible signal loss. The $t\bar{t}$ fraction remaining after this cut is approximately 30%.

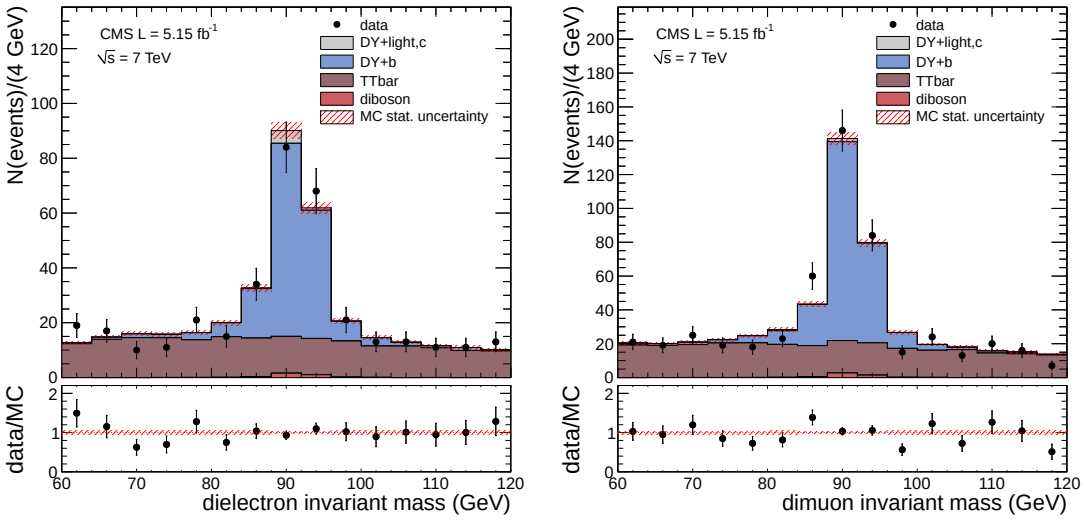

Figure 4.6: Dimuon (*left*) and dielectron (*right*) invariant mass in events with two reconstructed B candidates.

The Z boson candidate transverse momentum is shown in Fig. 4.7, for the dimuon and dielectron channels combined. The MC simulation reproduces reasonably well the Z spectrum.

As mentioned earlier in this chapter, the correct modelling of the properties of the secondary vertices and of the B-candidate kinematics is of great importance, as the reliability of the correction factors depends on it. The distributions of a set of relevant variables related to the SV reconstruction — the invariant mass, the track multiplicity, the three-dimensional flight distance significance — are shown in Fig. 4.8, for the leading and sub-leading B candidates. The simulation well reproduces the measured distributions for all variables.

The B-candidate kinematics is also studied. The transverse momentum of the leading and sub-leading B candidates and the p_T asymmetry, defined as the difference between leading and sub-leading B p_T divided by their sum, are shown in Fig. 4.9. The agreement in all these variables proves that not only the properties of the single B candidate are understood, but also the relations between the two particles.

Fig. 4.10 shows the detector-level distributions of the four angular variables of interest, with no cut on p_T^Z. Data are compared to the MC prediction, where the MG 4F simulation described in Section 1.3.5 provides for the signal component. A good agreement between data and MC is observed for the correlations between the Z boson momentum direction and the b-hadron system, while the data trends for two variables ΔR_{BB} and $\Delta\phi_{BB}$ are not well reproduced by the

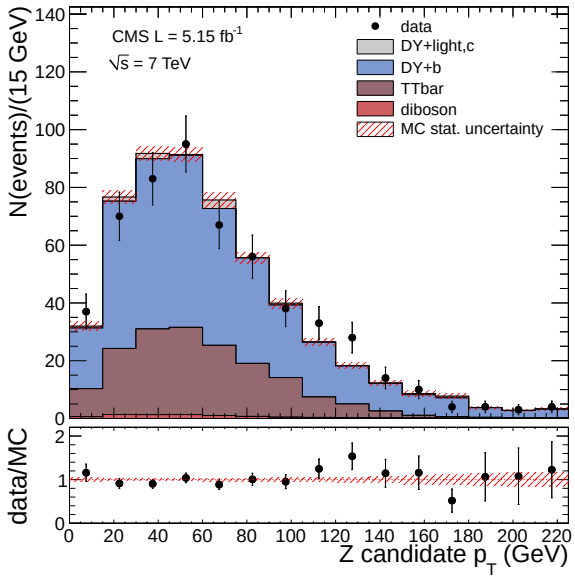

Figure 4.7: Z candidate transverse momentum, for the combination of the dimuon and dielectron final states.

simulation, the collinear B-hadron production being underestimated. A similar discrepancy is observed in the comparison between hadron-level distributions, and is extensively discussed in Section 4.5.

Fig. 4.11 shows the detector-level angles for the events passing $p_T^Z > 50$ GeV. The expected enhancement of the collinear b-hadron production cross section is visible is the ΔR_{BB} and $\Delta\phi_{BB}$ distributions. The comparison between data and the prediction leads to similar conclusions as for the inclusive sample.

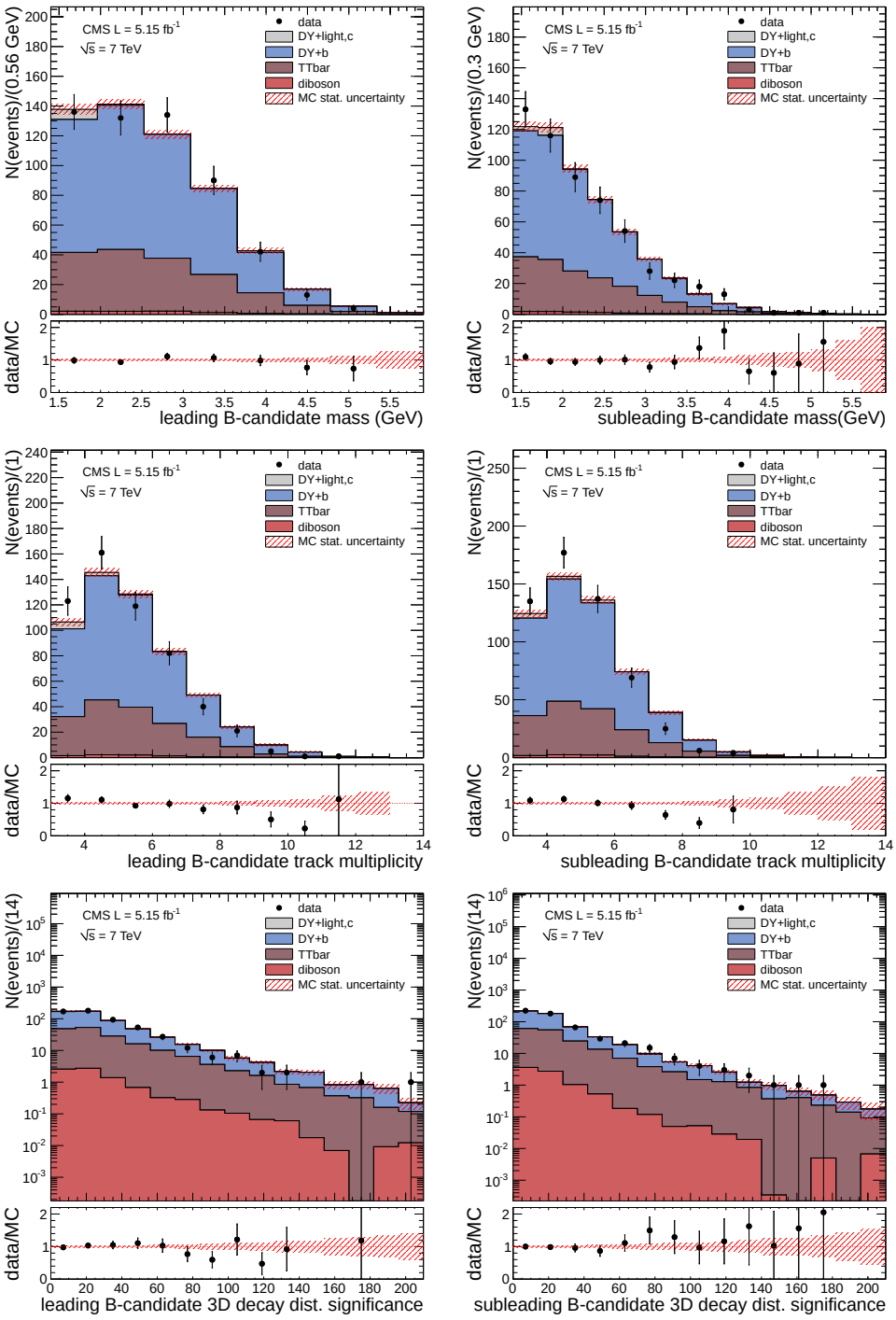

Figure 4.8: Properties of the leading (*left*) and sub-leading (*right*) reconstructed B candidates: invariant mass (*top*), track multiplicity (*middle*) and three-dimensional flight distance significance (*bottom*).

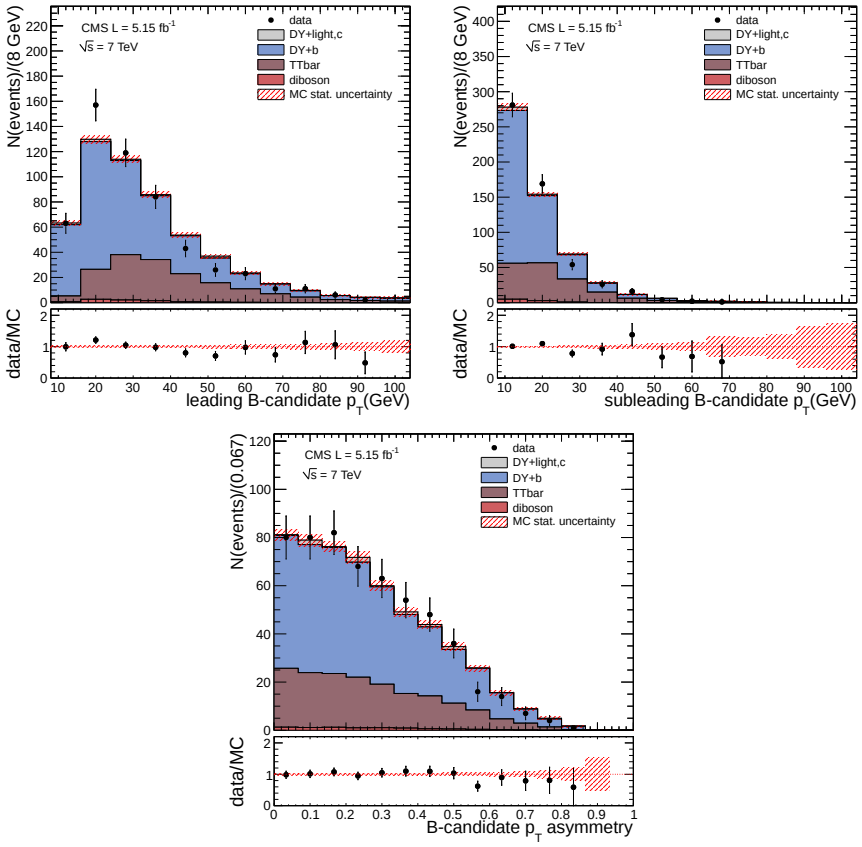

Figure 4.9: Transverse momentum of the leading (*top left*) and subleading (*top right*) reconstructed B candidates, and p_T asymmetry, as defined in the text.

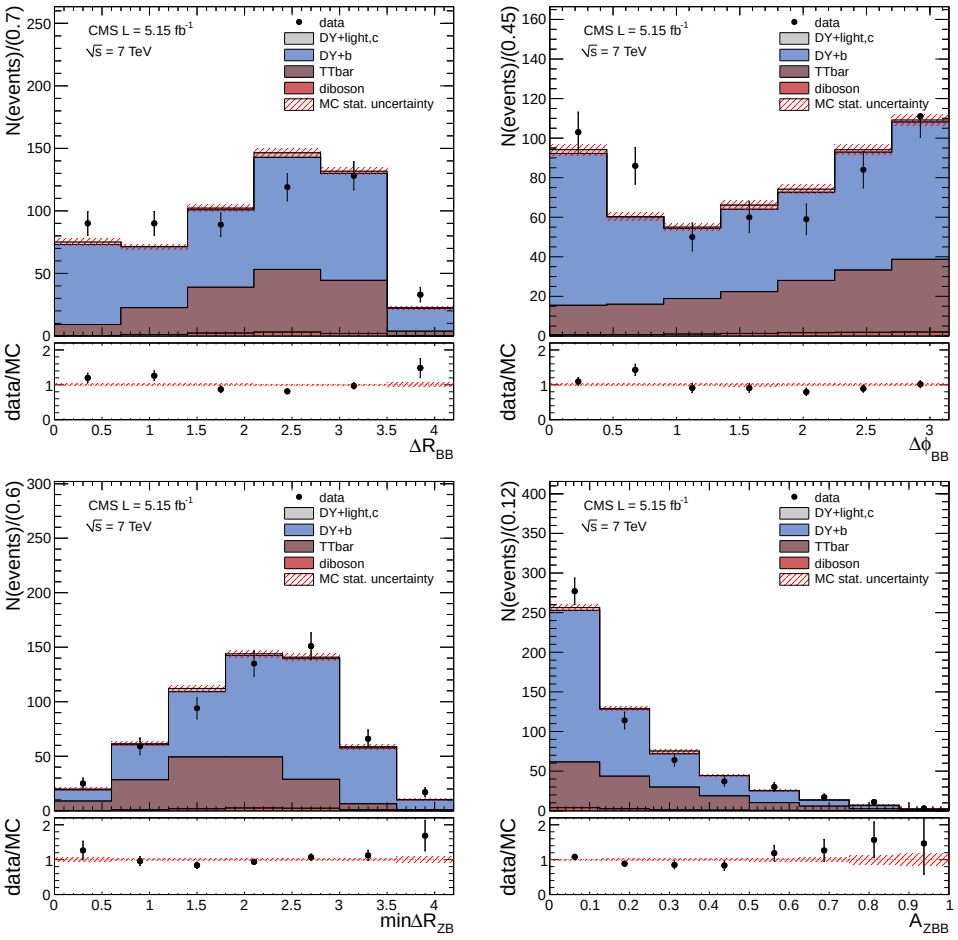

Figure 4.10: Detector-level distributions of the four angular variables of interest: ΔR_{BB}, $\Delta\phi_{BB}$, $min\Delta R_{ZB}$ and A_{ZBB}, with no cut on p_T^Z.

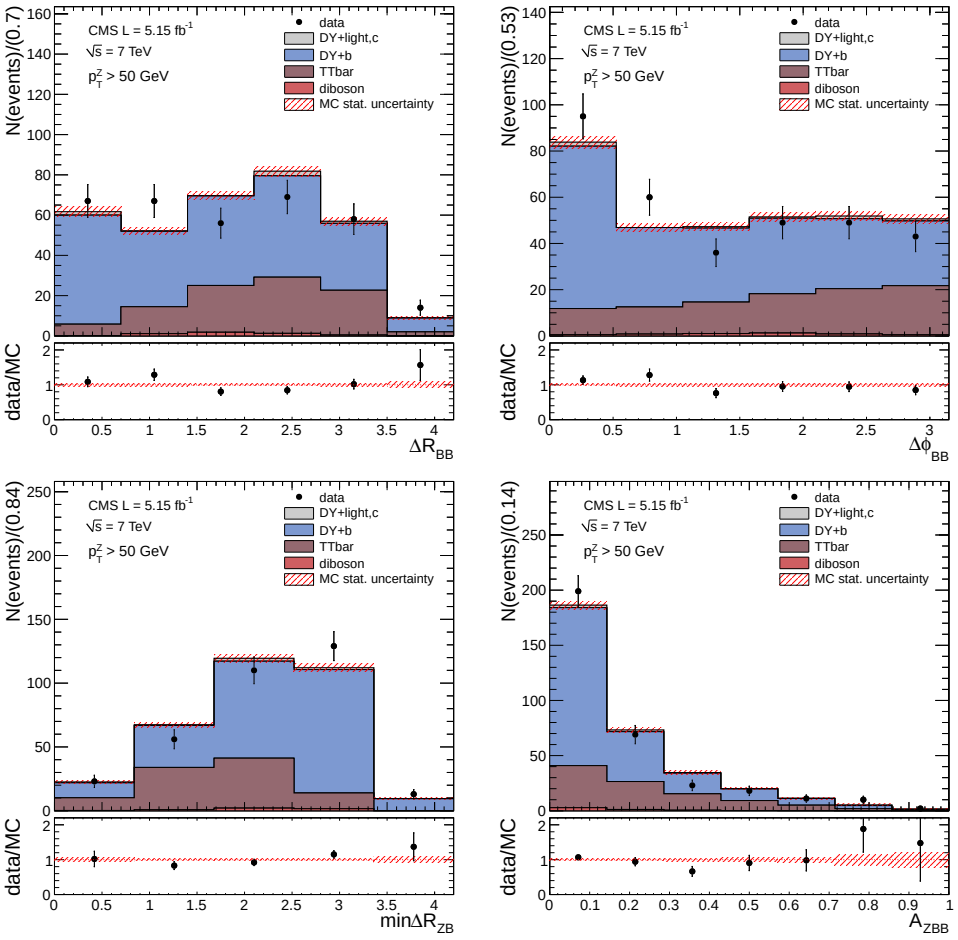

Figure 4.11: Detector-level distributions of the four angular variables of interest: ΔR_{BB}, $\Delta\phi_{BB}$, $min\Delta R_{ZB}$ and A_{ZBB}, with $p_T^Z > 50$ GeV.

4.3 Cross-section measurement

The core of the analysis described in this thesis is the measurement of the total and differential cross section of the process pp $\rightarrow Zb\bar{b}X$. As mentioned in the introduction, all cross sections are evaluated in the phase space defined by the following requirements:

- Lepton $p_T^\ell > 20$ GeV and $|\eta^\ell| < 2.4$;
- Dilepton invariant mass $81 < M_{\ell\ell} < 101$ GeV;
- b-hadron $p_T^B > 15$ GeV and $|\eta^B| < 2.0$.

The lepton momentum is corrected for final state radiation, i.e. the momentum of all the generator-level photons in a $\Delta R < 0.1$ cone around the lepton propagation direction is added to the lepton momentum. The b-hadron category includes any long-lived hadron containing one valence b-quark, and in particular the B^+, B^0, B_s, B_s^0, B_c^+ mesons and their antiparticles, and the Λ_b^0 baryon.

To obtain a hadron-level cross section, several effects have to be corrected for:

1. The $t\bar{t}$ background contamination, which constitutes approximately 30% of the sample of events passing all selection cuts. The Z signal yields N^{sig} are extracted with a fit of the dilepton invariant mass distribution, as discussed in Section 4.3.1. As differences in the mass distributions are expected between the two lepton flavours, the muon and electron yields are calculated independently;
2. The efficiency ϵ^ℓ ($\ell = \mu, e$) for reconstructing and selecting a lepton pair from a Z decay, and the acceptance $\mathcal{A}^\ell$. Also in this case, the correction factors for muons and electrons are calculated and applied separately. The combination of the two channels is consequently performed;
3. The efficiency ϵ^B for identifying a pair of b hadrons using the IVF technique;
4. The contamination from events containing reconstructed B candidates corresponding to mistagged charmed or light mesons, or to B hadrons outside the phase space, in particular with $p_T < 15$ GeV. The two components are treated independently, as they are affected by different systematic uncertainties. They are accounted for by the $\mathcal{P}$ and $\mathcal{S}^\mathcal{B}$ terms in Eq. 4.3;
5. The total integrated luminosity $\mathcal{L} = 5.15 \pm 0.11$ fb^{-1} of the data sample collected by CMS in 2011 and used for this analysis.

In general, the calculation of a differential cross section has to take into account the impact of the detector resolution, potentially responsible of bin-to-bin event migrations. For this

purpose, the correction factors are usually implemented as a matrix, where the diagonal entries account for the event reconstruction and selection efficiency and background contamination, while the off-diagonal terms correct for the bin-to-bin migrations. These can be approximately estimated by comparing the detector resolution with the width of the bins in the differential cross-section histograms. The ratio between the two gives an estimate of the fraction of events characterised by a true value of the variable of interest close enough to the bin edge, that the reconstructed value might fall into the adjacent bin as a pure effect of resolution. The angular resolution is discussed in Section 3.6 for the b-hadron related variables, ΔR_{BB} and $\Delta\phi_{BB}$, and in Section 4.3.6 for the correlations between the Z and b-hadron system. In all cases, it is between 20 and 30 times smaller than the bin size. In this analysis, therefore, the event migrations are considered negligible, and the efficiency and background corrections are estimated and applied independently in each bin of the differential cross-section histograms.

Summarising, the cross section is calculated using the formula:

$$\sigma_{i\alpha} = \mathcal{F}\left(\frac{N_{i\alpha}^{sig,\mu}}{\epsilon^{\mu}\cdot\mathcal{A}_{i\alpha}^{\mu}}, \frac{N_{i\alpha}^{sig,e}}{\epsilon^{e}\cdot\mathcal{A}_{i\alpha}^{e}}\right)\cdot\frac{\mathcal{S}_{i\alpha}^{B}}{\epsilon_{i\alpha}^{B}}\cdot\mathcal{P}_{i\alpha}\cdot\frac{1}{\mathcal{L}} \tag{4.3}$$

where the indices α and i respectively run over the angular variables and the bins. The function $\mathcal{F}$ represents the combination of the dilepton channels, consisting of a weighted average of the dimuon and dielectron yields, each corrected for the corresponding efficiency.

For brevity, this thesis reports figures illustrating the various steps of the analysis for the main angular variable, ΔR_{BB}. The event yields and the sets of corrections for the other variables are reported in the Appendix.

4.3.1 Signal extraction

Approximately one third of the sample of events passing all selection cuts is expected to be from $t\bar{t}$ production. A data-driven method is used to discriminate between the Z signal and the non-peaking $t\bar{t}$ background, consisting of an extended maximum likelihood fit of the dilepton invariant mass distribution. As some differences are expected in the signal and background shapes between the dimuon and dielectron final states — the Z boson peak being wider for electron decays due to FSR — the two channels are treated separately, and are combined in a later stage of the analysis. To eliminate any dependence of the hadron-level result on the MC description of the angular shapes, an independent fit is performed in each bin of the differential cross sections. For a better modelling of the background shapes, the wider invariant mass range $60 < M_{\ell\ell} < 150$ GeV is used in the whole fitting procedure.

The dilepton invariant mass distribution for the Z signal is parametrised using a Breigt-Wigner function $B(m, w)$ convolved with a Gaussian $G(\sigma)$. The Breigt-Wigner mean m and width w correspond to the Z boson mass and width respectively, while the Gaussian width accounts for the detector resolution. The Gaussian centroid is set to coincide to the Breigt-Wigner

mean. Conversely, there is no theoretical motivation guiding the choice of the $t\bar{t}$ background parametrisation. A third-degree Chebychev polynomial function $C(p_1, p_2, p_3)$ is used, being the one that provides the best description of the observed shapes. All parameters are free to float, and their values are extracted from two sequential unbinned maximum likelihood fits. To profit of the larger statistics and achieve a better precision, the entire sample of events with at least one reconstructed B candidate is used at this stage. This approach is justified by several tests showing no dependence of $M_{\ell\ell}$ shapes on the B-candidate multiplicity, as visible in Fig. 4.12.

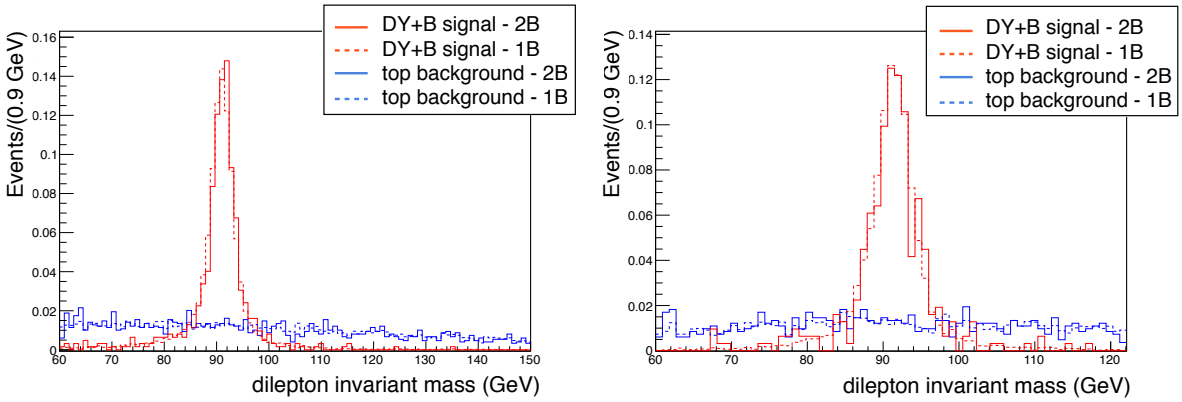

Figure 4.12: Dimuon (*left*) and dielectron (*right*) invariant mass distributions for simulated $Zb\bar{b}$ signal and $t\bar{t}$ background events with one or two reconstructed B candidates. The mass shapes in the two samples with different B-candidate multiplicity are in good agreement.

The background parameters are extracted from a maximum likelihood fit of the $t\bar{t}$ MC sample, and fixed. The $t\bar{t}$ process is well understood [69,70], and the simulation is highly reliable. The Chebychev polynomial is added to the signal PDF, and the measured $M_{\ell\ell}$ distribution is fit with the combined function to determine the Breigt-Wigner and Gaussian parameter values. The resulting PDF is:

$$f(m, w, \sigma, p_1, p_2, p_3) = B(m, w) \otimes G(\sigma) + C(p_1, p_2, p_3) \tag{4.4}$$

As the invariant mass shapes are expected to depend on the kinematics, a different set of parameters is extracted for each p_T^Z requirement. The values and errors of the signal parameters are listed in Table 4.4. Fig. 4.13 and 4.14 show the background and signal+background distributions from MC and data respectively, fitted with the functions described above, for the inclusive p_T^Z spectrum and the $p_T^Z > 50$ GeV cut as examples. Similar results are found for the other p_T^Z ranges.

Once that all parameters are determined, their values and their uncertainties are fixed. The

	dimuon			dielectron		
p_T^Z cut (GeV)	m (GeV)	w (GeV)	σ (GeV)	m (GeV)	w (GeV)	σ (GeV)
inclusive	90.82 ± 0.03	3.36 ± 0.14	0.94 ± 0.10	91.10 ± 0.05	3.64 ± 0.21	1.76 ± 0.12
40	90.82 ± 0.05	3.67 ± 0.21	0.76 ± 0.19	91.24 ± 0.07	3.83 ± 0.31	1.50 ± 0.17
50	90.83 ± 0.05	3.85 ± 0.29	0.71 ± 0.27	91.33 ± 0.07	3.75 ± 0.31	1.51 ± 0.18
80	90.90 ± 0.09	4.19 ± 0.67	0.59 ± 0.69	91.40 ± 0.11	3.30 ± 0.51	1.49 ± 0.29
120	90.91 ± 0.15	4.15 ± 0.69	0.78 ± 0.74	91.22 ± 0.18	3.71 ± 0.70	1.26 ± 0.47

Table 4.4: Shape parameters for the signal Breit-Wigner and Gaussian functions, for the dimuon and dielectron final state.

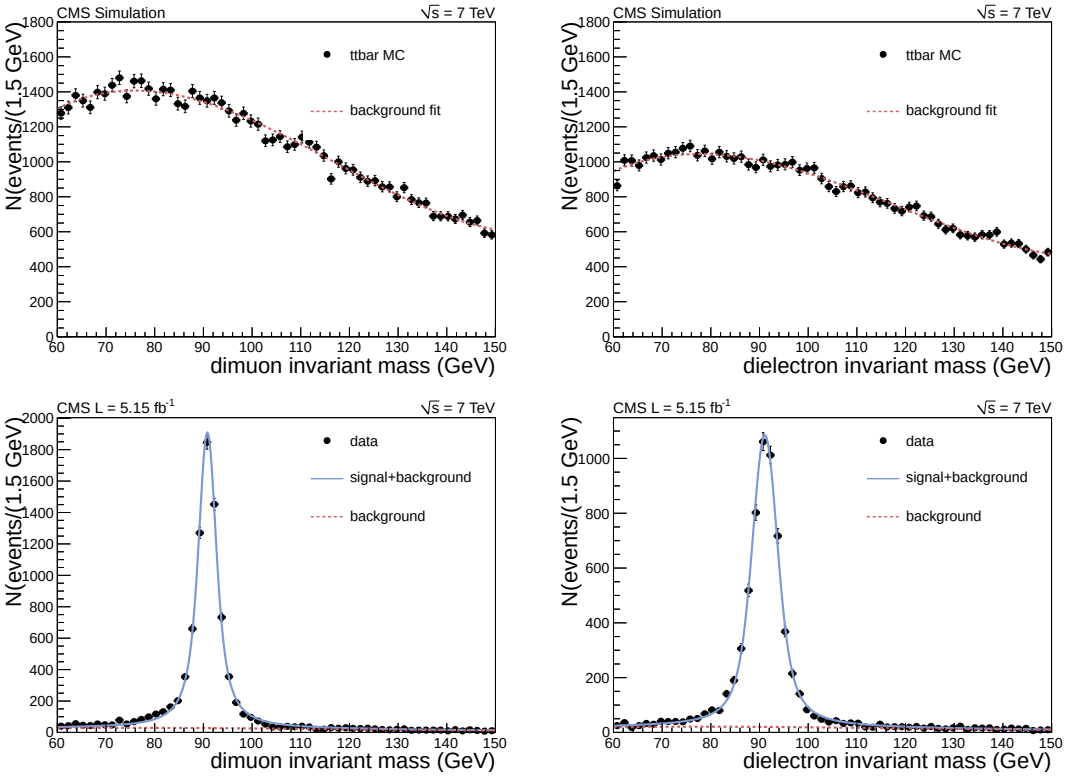

Figure 4.13: Dimuon (*left*) and dielectron (*right*) invariant mass distributions for the $t\bar{t}$ background (*top*) and for the signal+background (*bottom*), fitted with the functions described in the text. The background shapes are extracted from the MC, while the signal+background fit is performed on data. No p_T^Z cut is applied.

fitting functions are extended by Poisson terms to account for the total number of signal (N^{sig}) and background (N^{bckg}) events in the $60 < M_{\ell\ell} < 150$ GeV range. N^{sig} and N^{bckg} are the

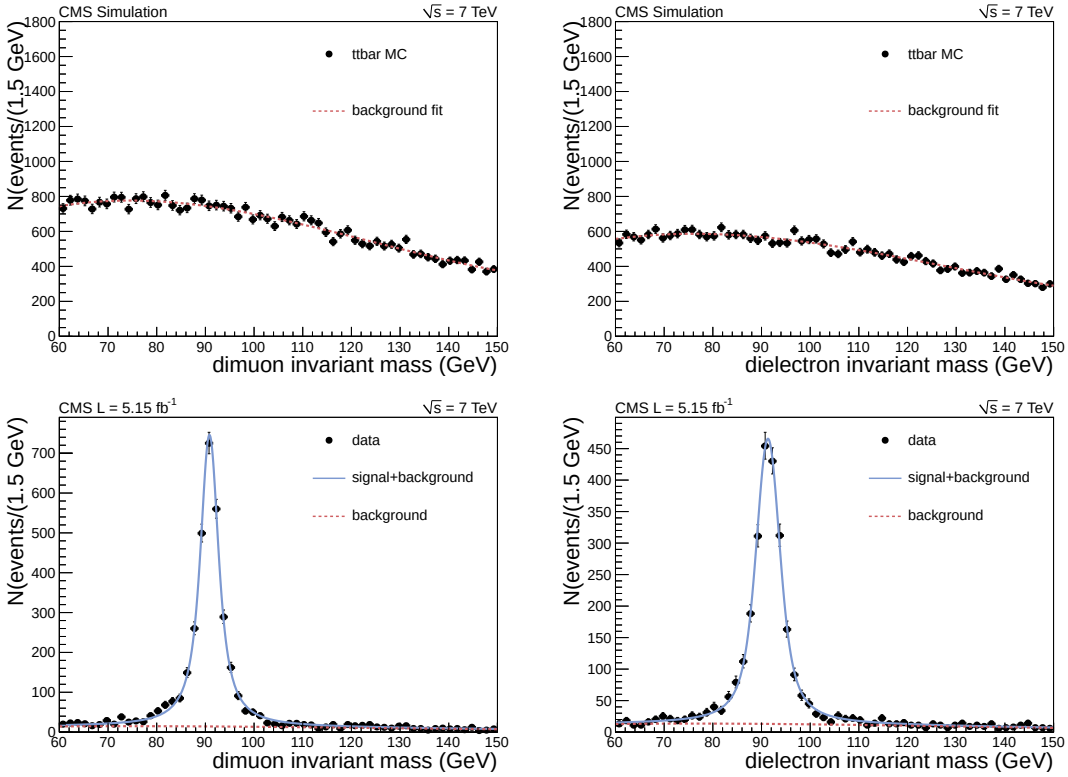

Figure 4.14: Dimuon (*left*) and dielectron (*right*) invariant mass distributions for the $t\bar{t}$ background (*top*) and for the signal+background (*bottom*), fitted with the functions described in the text. The background shapes are extracted from the MC, while the signal+background fit is performed on data. The $p_T^Z > 50$ GeV cut is applied.

only free parameters in the unbinned maximum likelihood fit applied to the sample of events passing the final selection. For the differential cross-section measurements, an independent fit is performed in each bin of the angular variables, while for the total cross-section estimation the entire event sample is used. The sum of the yields obtained in the single bins gives however a consistent result. Fig. 4.15 shows the final fit of the $M_{\ell\ell}$ distribution in the dimuon and dielectron channels, for the total cross-section calculation. Similar results are obtained in all bins of the differential cross sections. All fit results are shown in Appendix A.1. The signal yields are subsequently extrapolated to the narrower window $81 < M_{\ell\ell} < 101$ GeV. The resulting signal yields as a function of ΔR_{BB} are shown in Fig. 4.16. Only the statistical errors extracted from the likelihood fit are reported.

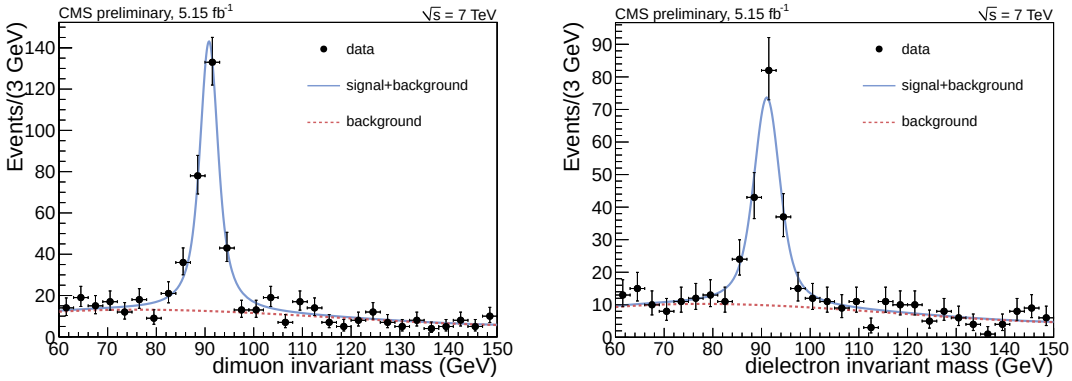

Figure 4.15: Dimuon (*left*) and dielectron (*right*) invariant mass distributions from data, fitted with the functions described in the text, for the signal yield extraction. No p_T^Z cut is applied.

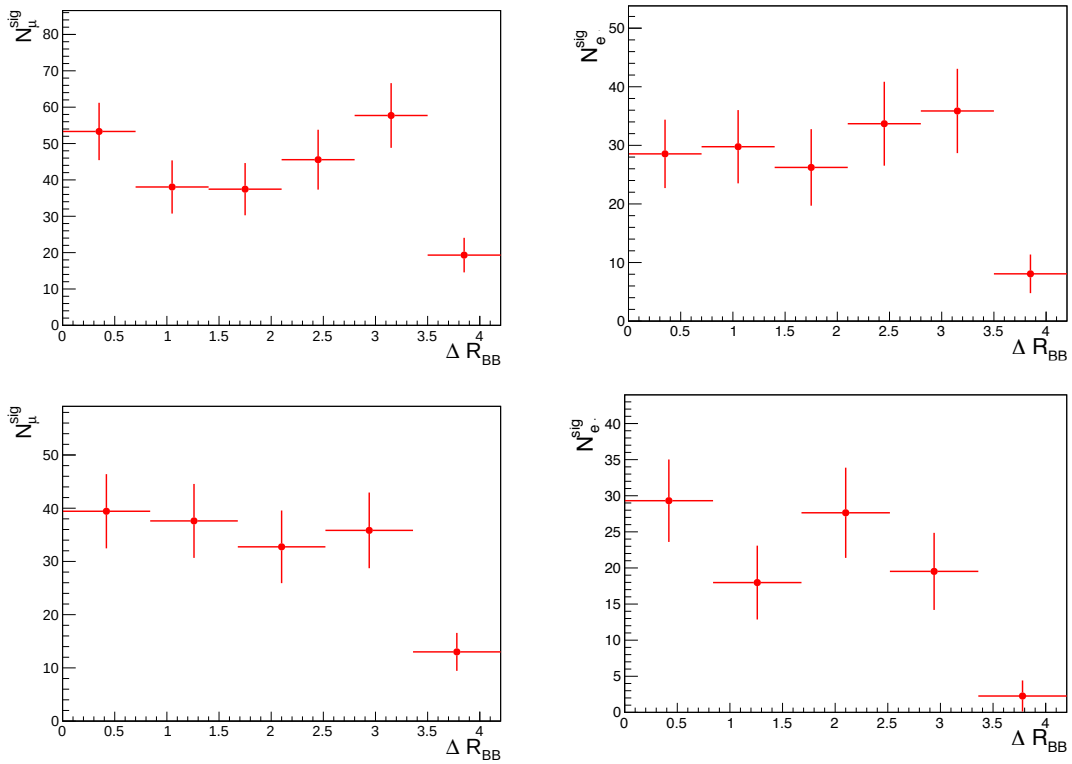

Figure 4.16: Dimuon (*left*) and dielectron (*right*) signal yields as a function of ΔR_{BB}, for the inclusive p_T^Z event sample (*top*), and for the $p_T^Z > 50$ GeV cut (*bottom*).

Closure test

A closure test based on the MC simulation is performed, to exclude any potential bias in the extraction of the signal and background yields. A known number of simulated signal and $t\bar{t}$ background events is combined into a single sample, which approximately reproduces the fractions expected in data. To profit of a higher statistics, the fitting procedure is applied to the entire sample, as for the calculation of the total cross section, and the output yields are compared to the true values. The results are collected in Table 4.5, showing an excellent agreement between the returned values and the input. The test is repeated varying the signal and background fractions. No significant bias is observed.

	dimuon		dielectron	
	simulated	fit	simulated	fit
signal	614	602 ± 29 (stat.) ± 7 (syst.)	480	494 ± 26 (stat.) ± 6 (syst.)
background	223	230 ± 9 (stat.) ± 7 (syst.)	178	163 ± 8 (stat.) ± 7 (syst.)

Table 4.5: For the dimuon and dielectron channels, signal and background yields returned by the fit performed on a simulated dataset combining $t\bar{t}$ and $Zb\bar{b}$ events. The fit results are compared to the known number of signal and background events. No p_T^Z cut is applied.

4.3.2 Dilepton selection efficiency and acceptance

Before being combined as in Eq. 4.3, the signal yields of the dimuon and dielectron channels have to be independently corrected for the respective lepton reconstruction and selection efficiencies (ϵ^ℓ) and for a factor ($\mathcal{A}^\ell$) accounting for the detector acceptance. The lepton efficiency and acceptance correction is decomposed into two terms:

$$(\epsilon^\ell \cdot \mathcal{A}^\ell) = (\epsilon^\ell_{\text{trigger}})_{i\alpha} \cdot (\epsilon^\ell_{\text{offline}} \cdot \mathcal{A}^\ell)_{i\alpha} \tag{4.5}$$

where $\epsilon^\ell_{\text{trigger}}$ is the efficiency of the trigger paths for the specific event topology, and $\epsilon^\ell_{\text{offline}}$ is the efficiency of the offline selection cuts listed and discussed in Section 4.1, and the indices i and α run over the bins and the angular variables. The acceptance term $\mathcal{A}^\ell$ extrapolates the yields measured in the experimental fiducial phase space to the phase space chosen for the cross-section measurement. For the dimuon final state, the two spaces coincide, and $\mathcal{A}^\ell$ is equal to one. For the dielectron channel, $\mathcal{A}^\ell$ accounts for the $1.442 < |\eta| < 1.566$ cut aimed at excluding the non-instrumented region between ECAL barrel and endcap. This correction is estimated as a function of the angular variables together with $\epsilon^\ell_{offline}$.

Both data and the MC simulation are used to extract the $\epsilon^{\ell}_{\text{trigger}}$ and $\epsilon^{\ell}_{\text{offline}} \cdot \mathcal{A}^{\ell}$ terms. In general, the measurements based on data are performed with the so-called tag & probe approach, which evaluates lepton reconstruction efficiencies exploiting known resonances. For the lepton p_T ranges covered by this analysis, the Z boson is the most appropriate object. Pairs of same flavour leptons are considered: one of them — the *tag* — is required to pass tight cuts, in order to ensure high purity, while the other one — the *probe* — is used to measure the efficiency provided by the selection criteria of interest. The lepton pairs are combined into a Z candidate, and an extended maximum likelihood fit is applied to their invariant mass distribution to determine the total number of Z decay events, and the number of Z candidate events where the probe passes the selection cuts. A Crystal Ball function and an exponential are chosen to model the signal and background shapes. The selection efficiency is finally estimated as the ratio between the two yields.

Trigger efficiency

Since the emulation of the dilepton trigger selections adopted in this analysis is not implemented in the simulation, the $\epsilon^{\ell}_{\text{trigger}}$ term can only be measured from data with the tag & probe technique [68]. $\epsilon^{\ell}_{\text{trigger}}$ is defined as the probability that a selected lepton passes the HLT filters. In order to avoid any bias, it has to be estimated on samples of events filtered by trigger selections different from those adopted. Single muon and electron triggers with loose requirements, namely low p_T thresholds and relaxed isolation cuts [68], are used for that purpose. Only events with two reconstructed leptons passing the analysis selection cuts are retained. The efficiency is evaluated as the ratio between the number of events where the probe is geometrically matched to the object that fired the double lepton trigger, and the total number of selected events, evaluated with a maximum likelihood fit of the dilepton invariant mass distribution. The measurement is performed in bins of p_T and η of the two legs.

The main systematic uncertainty on $\epsilon^{\ell}_{\text{trigger}}$ is assumed to originate from the fit, and in particular from the parametrization of the signal and background invariant mass shapes.

The tag & probe procedure described above is repeated for all the unprescaled dimuon and dielectron trigger paths used in the different run periods. The resulting efficiency values for the single leg are collected in [68].

The efficiencies of the dimuon and dielectron trigger paths adopted in the different data taking periods are combined into a single number, estimated as the average of the single path efficiencies, each weighted by the integrated luminosity collected in the corresponding run period. Typical dilepton trigger efficiency values are around 96% for the dimuon and 98% for the electron channel.

Offline selection efficiency and combination

The offline reconstruction and selection efficiency $\epsilon^{\ell}_{\text{offline}}$ is estimated using both data and the simulation.

The measurement from data exploits the tag-and-probe method described earlier in this chapter [68]. The analysis uses a sample of events containing a Z boson and at least one jet with $p_T > 20$ GeV and $|\eta| < 2.1$. Since a large dataset is needed for the evaluation of the efficiencies in bins of the lepton kinematics, no requirement is imposed on the jet flavour, which is assumed to be irrelevant. The tag & probe procedure is applied to data and simulation and the results are found to be in good agreement, the discrepancies being of the order of few per cent. The ratio between data and the prediction is interpreted as p_T and η-dependent scale factor, and used to rescale the MC efficiency values before applying them to data. The complete set of scale factors is documented in [68].

The angular correlation measurements require an estimation of the dilepton efficiency and acceptance corrections as a function of the measured angles. Due to the limited size of the data sample, the dependence on the correlation variables is investigated using the MC simulation. The second term in Eq. 4.5 ($\epsilon^{\ell}_{\text{offline}} \cdot \mathcal{A}^{\ell}$) is calculated with the following formula:

$$(\epsilon^{\ell}_{\text{offline}} \cdot \mathcal{A}^{\ell})_{i\alpha} = \left[\frac{N(2\ \ell^{\text{reco}}_{\text{matched}}\ \textit{and}\ 2\ B^{\text{gen}}_{\text{accept}})}{N(2\ \ell^{\text{gen}}_{\text{accept}}\ \textit{and}\ 2\ B^{\text{gen}}_{\text{accept}})} \right]_{i\alpha}. \tag{4.6}$$

The number of events with two generated b hadrons and two generated leptons within the acceptance constitutes the denominator. The generated leptons are defined by combining the lepton momentum with the energy of all the photons in a $\Delta R < 0.1$ cone surrounding the lepton propagation direction. This approach permits to recover the energy emitted as final state radiation, and to estimate the original momentum of the hard scattering product. The numerator in Eq. 4.6 is given by the number of events where each of the generated leptons is matched to a reconstructed and selected one, according to the geometrical $\Delta R < 0.1$ criterion. The calculation is performed independently for each of the p^Z_T cuts. The resulting angle-dependent correction factors are rescaled on an event-by-event basis to match the values observed in data, as described in the next subsection.

Combination of dilepton efficiency factors

The extended maximum likelihood fit described in Section 4.3.1 discriminates the signal and background components on a statistical basis. As the event-by-event information about the lepton kinematic properties is not available after the signal extraction, the p_T and η-dependent corrections for the data/MC offline efficiency scale factors and the trigger efficiencies have to be embedded in the computation of the bin-by-bin dilepton efficiency corrections on MC. For this purpose, each simulated event in the numerator of Eq. 4.6 is reweighted by a coefficient equal

to the product of the trigger and offline efficiency rescaling factors. The resulting efficiency corrections are shown in Fig. 4.17 for the ΔR_{BB} variable, as example. Only the statistical uncertainties are reported.

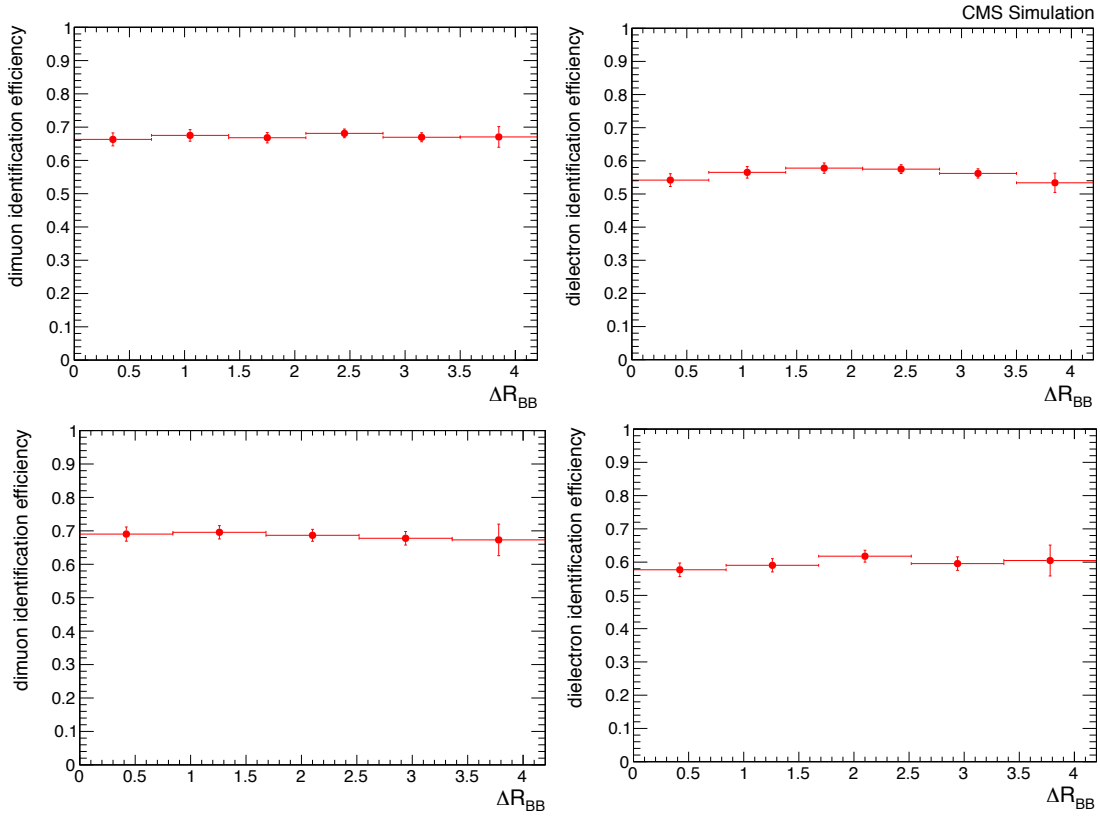

Figure 4.17: Dimuon (*left*) and dielectron (*right*) trigger and offline selection efficiency and acceptance corrections as a function of ΔR_{BB}, for the inclusive p_T^Z spectrum (*top*) and for the $p_T^Z > 50$ GeV cut (*bottom*). The trigger efficiency correction and the data/MC scale factors are applied on an event-by-event basis. Only statistical errors are shown.

4.3.3 Combination of dimuon and dielectron channels

As described in the previous sections, the dimuon and dielectron signal yields are extracted separately, and are independently corrected for the dilepton reconstruction and selection efficiencies. The two final states are combined before applying the corrections related to the b-hadron identification, common to both channels.

The combination is performed using the BLUE package [71, 72], which sums up the dimuon

and dielectron contributions reweighted according to the respective statistical and systematic errors. The correlations among the different sources of uncertainty — the statistical and systematic errors on N^{sig} from the extended maximum likelihood fit, the systematic error on ϵ^{ℓ} — are assigned as prior assumption to the algorithm. Several correlation hypothesis are tested: the combination procedure is repeated under variation of the correlation coefficients between 0 and 1, in steps of 0.1. The effect of the correlation assumption on the result is negligible. BLUE provides the statistical and systematic uncertainties associated to the combination.

4.3.4 b-hadron pair identification efficiency

The b-hadron pair identification and selection efficiency (ϵ^{B}) is extracted from the simulation, using a sample of events containing two reconstructed leptons matched to the generated ones from the Z boson decay. It is calculated as a ratio between the number of events with exactly two generated b hadrons within the acceptance matched to reconstructed B candidates passing all selection cuts, and the total number of events with exactly two generated b hadrons within the acceptance:

$$(\epsilon^{\mathrm{B}})_{i\alpha} = \left[\frac{N(2\ \ell^{\mathrm{reco}}_{\mathrm{matched}}\ and\ 2\ \mathrm{B}^{\mathrm{reco}}_{\mathrm{matched}})}{N(2\ \ell^{\mathrm{reco}}_{\mathrm{matched}}\ and\ 2\ \mathrm{B}^{\mathrm{gen}}_{\mathrm{accept}})} \right]_{i\alpha} \tag{4.7}$$

The generated b hadron is matched to a reconstructed B candidate if the three-dimensional angle ΔR between their flight directions is smaller than 0.1, five times the angular resolution provided by the IVF. The hadron-level cross-section results are proved to be robust with respect to the choice of the matching ΔR value. Eq. 4.7 is independently applied in each cross-section bin, and leads to efficiency values between 8 and 13%. Fig. 4.18 and Table 4.6 show the ϵ^{B} correction factors for the extraction of the ΔR_{BB} differential cross section, and for the total cross-section estimation. A finer binning is used in the plots compared to the final cross-section measurement, to give a more detailed description of the efficiency trend. Significant variations are observed among the different cross-section bins and the various p_T^Z regions. These fluctuations are mostly motivated by a strong dependence of the secondary vertex reconstruction and B-candidate selection efficiency on the decaying particle kinematics, in particular on the transverse momentum. A combination of two different effects, instead, causes the efficiency drop observed in the $\Delta R_{BB} < 0.2$ range: at such low angular separations, the two displaced decays start overlapping, and the assignment of tracks to the correct vertex is more difficult, resulting in low-quality reconstructed SV. Furthermore, the two collinear SV can be misidentified as a sequential b $\rightarrow$ c hadron decay and merged into a single B candidate.

A global rescaling factor is applied to the efficiency curves extracted from the simulation, correcting for the discrepancy between data and MC simulation observed in the p_T^{rel} analysis described in Section 3.6. The data/MC disagreement can be either related to a mismodelling of the b-hadron production mechanisms — leading to inconsistencies in the event kinematics — or to an imperfect description of the details of tracking and secondary vertex reconstruc-

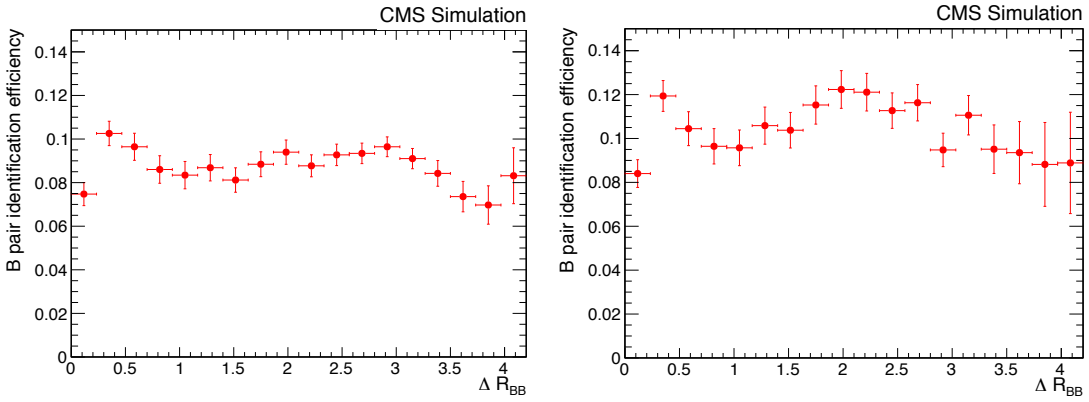

Figure 4.18: Efficiency for identifying and selection a b-hadron pair as a function of ΔR_{BB}, with no cut on p_T^Z (*left*) and for $p_T^Z > 50$ GeV (*right*), as measured in the MC simulation. Only the statistical uncertainties are shown.

p_T^Z cut (GeV)	ϵ^B ($\times 10^{-2}$)
0	8.88 ± 0.13
40	10.02 ± 0.18
80	11.87 ± 0.31
120	12.55 ± 0.48

Table 4.6: Efficiency corrections for the total cross-section estimation, as a function of the minimum p_T^Z cut. Only the statistical uncertainties are shown.

tion. The first category is physics-process dependent, and should therefore be investigated on the event topology of interest. Such a study is unfortunately not possible in the context of this analysis, due to the small size of the selected data sample. Although extracted from an inclusive jet event sample, the data/MC rescaling factors are however evaluated in bins of the jet p_T. The dependence on the event kinematics is therefore factorized, and the resulting scale factors only account for effects related to the SV reconstruction itself.

As shown in Fig. 3.19, while the efficiency values depend on the jet transverse momentum, the data/MC discrepancy exhibits no significant variations in the considered range, resulting in an approximately constant scale factor. The application of a constant rescaling of the entire efficiency curves is therefore sufficient. This approach considerably simplifies the analysis, given the technical difficulty of the conversion between the p_T of the b-jet — which reproduces well the p_T of the b parton — and the p_T of the reconstructed B candidate, due to the poor momentum resolution provided by the IVF.

4.3.5 b-hadron purity and soft b-hadron contamination

The maximum likelihood fit described in Section 4.3.1 estimates the backgrounds characterized by the absence of a Z boson, highly dominated by the $t\bar{t}$ production. Nevertheless, additional sources of background have to be considered. In particular, three event categories can affect the total cross-section estimation:

- $pp \rightarrow Zc\bar{c}X$. The IVF algorithm is used in the configuration that provides the highest purity, and the lowest efficiency for identifying charmed hadrons; The probability that both charm hadrons are mistagged is therefore substantially negligible, leading to a contamination of approximately 1-2%;
- $pp \rightarrow Zb\bar{b}b\bar{b} + X$. The fraction of events with an additional b-hadron pair is determined by the probability of a gluon splitting into a $b\bar{b}$ pair, which is of the order of 0.1%. Nevertheless, the efficiency for tagging at least two b hadrons is higher than for events with a single b-hadron pair, resulting in a 2-3% contribution to the selected event sample;
- $pp \rightarrow Zb\bar{b}X$, with at least one of the two b hadrons outside the selected phase space, and in particular below the p_T threshold. This contamination is due to the poor momentum resolution provided by the IVF, for which there is a significant probability that a soft b hadron decaying mostly into charged daughters leads to a B candidate passing the selection requirements. This component is referred to as *soft b-hadron contamination*.

Another type of background consists of signal $Zb\bar{b}$ events, in which at least one of the b hadrons is not reconstructed, while a charmed hadron is misidentified as a B candidate. Although such configurations do not affect the estimation of the total cross-section normalisation, as they constitute true $Zb\bar{b}$ events, they are responsible for migrations of events in the angular correlations, and need to be accounted for. Two distinct sources are considered within this category:

- $pp \rightarrow Zb\bar{b}cX$, where the charmed particles originate from initial state radiation and fragmentation. From the simulation, this process gives the largest contribution to the non-b contamination;
- $pp \rightarrow Zb\bar{b}X$, with a sequential $b \rightarrow c$ decay. The B-candidate producer described in Section 3.6 identifies secondary vertex pairs from such processes, and merges them into a single object. If this procedure fails, the charmed hadron is tagged as a separate B candidate, resulting in a wrong estimation of the correlation angles, as shown in Fig. 4.19. Together with the previous term, this is the dominant source of impurity.

As mentioned in the introduction to this chapter, the different Z-background categories listed above are grouped into two main correction factors: the B purity and the soft B contamination. More details about their estimation and their uncertainties are given in the next two sections.

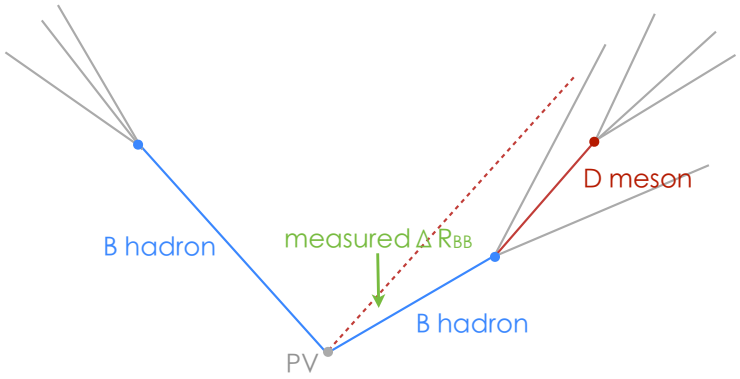

Figure 4.19: Reconstruction of a sequential b→c decay, in which the b and the charmed hadrons are not correctly merged into a single B candidate. This configuration results in a biased estimate of the angular separation.

B purity

The B hadron purity correction ($\mathcal{P}$) is evaluated from the MC simulation. For b-jet tagging algorithms, a data-driven method is generally available for the measurement of $\mathcal{P}$, consisting of a maximum likelihood fit of the secondary vertex invariant mass or flight distance significance distributions. These variables allow in fact to distinguish the true b-hadron decays from the misidentified light and charmed particles. Nevertheless, all variables with such discrimination power are included in the IVF B-candidate selection, and are therefore affected by an intrinsic bias. Moreover, the charm contamination is low — approximately 12% — and the size of the data sample is not sufficient for a precise determination of its fraction. Hence, the data-driven technique is unfortunately not applicable in this analysis, which has to rely on the simulation.

The MC-based calculation uses a simulated sample of events with a pair of reconstructed muons or electrons passing all selection cuts listed in Section 4.1.2, matched to the generated leptons from the Z decay. The purity is estimated as a ratio between the number of events with two reconstructed B candidates passing all selection cuts and matched to two generated b hadrons, and the total number of events with two reconstructed B candidates passing all selection cuts, as shown in Eq. 4.8.

$$\mathcal{P}_{i\alpha} = \left[\frac{N(2\ \ell^{\text{reco}}_{\text{matched}}\ \textit{and}\ 2\ \text{B}^{\text{reco}}_{\text{matched}})}{N(2\ \ell^{\text{matched}}_{\text{reco}}\ \textit{and}\ 2\ \text{B}^{\text{reco}}_{\text{selected}})} \right]_{i\alpha} \tag{4.8}$$

Reconstructed and simulated objects are matched if the three-dimensional angular distance ΔR between their flight directions is smaller than 0.1. As for the efficiency, the results are proved not to be affected by the choice of the matching ΔR value. In order not to interfere with the soft b-hadron contamination term, no acceptance requirement is applied to the b hadrons at the generator level.

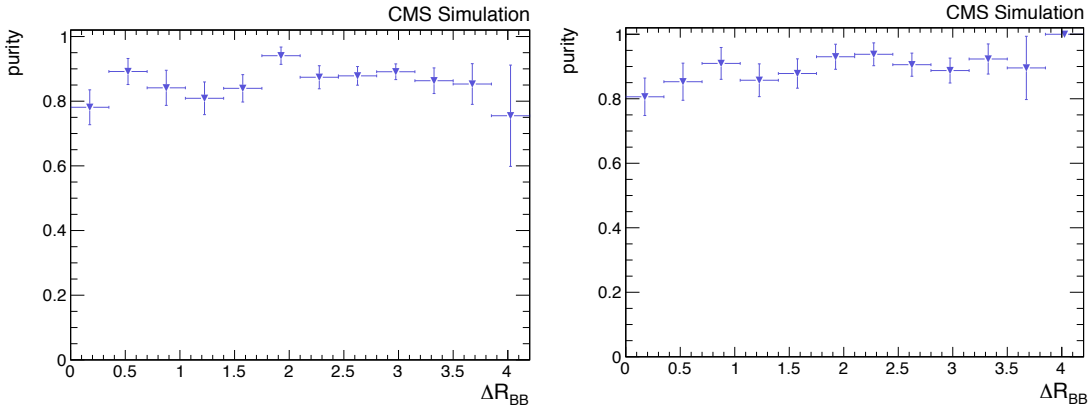

Figure 4.20: B purity as a function of ΔR_{BB}, with no cut on p_T^Z (*left*) and for $p_T^Z > 50$ GeV (*right*). Only the statistical errors are shown.

p_T^Z cut (GeV)	$\mathcal{P}$ ($\times 10^{-2}$)	$\mathcal{S}$ ($\times 10^{-2}$)
0	93.8±0.8	86.4±1.2
40	94.7±0.9	87.9±1.3
80	89.8±1.9	96.9±1.1
120	86.0±3.2	95.7±1.8

Table 4.7: B purity and soft b-hadron corrections for the total cross-section estimation, for the different p_T^Z cuts. Only the statistical errors are shown.

Fig. 4.20 and Table 4.7 show the B purity corrections for the extraction of the hadron-level total cross section and of the angular correlation distributions. The differential purity correction plot has a finer binning compared to the cross-section histogram, to give a better description of the trend at low ΔR_{BB}. As expected, the purity is substantially constant as a function of the angular separation variable. The only visible feature is a decrease for $\Delta R_{BB} < 0.2$, which has the same origin of the drop in efficiency in the same region of the angular spectrum.

Soft b-hadron contamination

The correction factor accounting for the contamination from $Zb\bar{b}$ events with at least one b hadron outside the acceptance region is estimated with the same simulated sample used for the purity measurement. It is defined as $1 - f$, where f is the fraction of events with at least one reconstructed B candidate passing all selection requirements, matched to a generated b hadron with $p_T < 15$ GeV. The matching is performed according the ΔR criterion described in the previous sections. The resulting correction factors for the differential ΔR_{BB} distribution

and for the total cross-section estimation are shown in Fig. 4.21 and Table 4.7. The other results are collected in Appendix A.

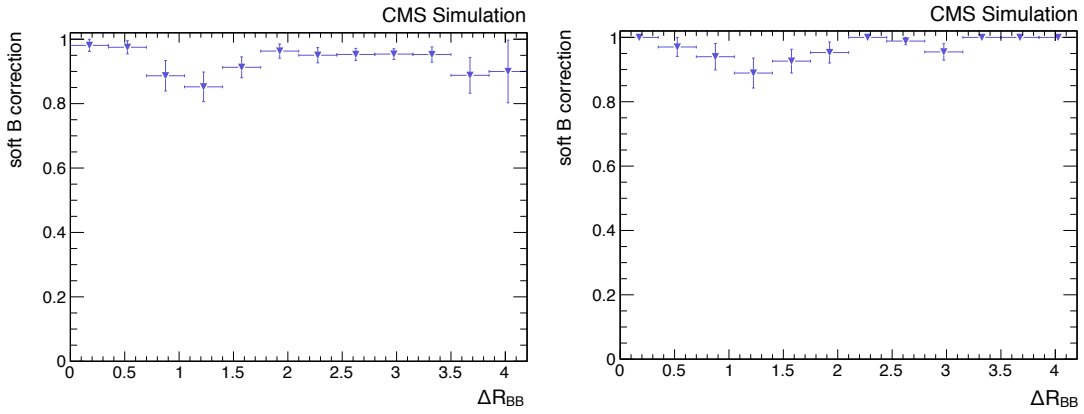

Figure 4.21: Soft B hadron corrections as a function of ΔR_{BB}, for the inclusive p_T^Z spectrum (*left*) and for the $p_T^Z > 50$ GeV cut (*right*). Only the statistical errors are shown. A finer binning is used compared to the final cross-section histogram, to give a more detailed description of the trend.

4.3.6 Statistical and systematic uncertainties

The statistical uncertainty on the measured cross section is estimated by the maximum likelihood fit used to determine the dimuon and dielectron signal yields, and is between few % and 15% for the total cross section, and around 10% for the most populated bins of the differential cross sections.

Furthermore, systematic uncertainties are assigned to each of the corrections in Eq. 4.3, as discussed below.

Extraction of dimuon and dielectron signal yields

The main systematic uncertainty related to the extraction of the signal yields is due to the dilepton invariant mass shape parametrization, for the extended maximum likelihood fit. It is estimated by performing the fit several times varying the fitting function parameters within a given range. For the signal function, this range is chosen to be the uncertainty on the parameters themselves, given by the preliminary fit of the single B candidate event sample. For the background, instead, since the procedure relies upon the goodness of the MC simulation, a conservative 20% is considered. The maximal variation of the signal yields among all bins of the differential cross sections, approximately equal to 2%, is taken as systematic uncertainty.

This error is propagated to through the combination of the dimuon and dielectron channels, to the final cross section calculation.

Dilepton reconstruction and selection efficiency, dimuon and dielectron channel combination

The main systematic uncertainty affecting the ϵ^{ℓ} term is related to the measurement of the dilepton trigger efficiency and of the offline efficiency scale factors. As mentioned above, the uncertainties on the rescaling factors are estimated as a function of the lepton kinematics, either by varying the parametrization of signal and background shapes in the maximum likelihood fit, or by comparing the results obtained with two alternative methods. These uncertainties are propagated to the ϵ^{ℓ} correction term. For this purpose, ϵ^{ℓ} is re-assessed under variation of the two rescaling factors within their errors, the maximal variation in all angular variable bins being approximately 1% for both dimuon and dielectron final states.

The systematic uncertainties on the signal yields and on the dilepton efficiency corrections are given as input to the BLUE algorithm described in Section 4.3.3, performing the combination of dimuon and dielectron channels. The output systematic uncertainty on the resulting corrected yield is approximately 2%.

b-hadron pair identification efficiency

As the b-hadron identification and selection efficiency corrections are extracted from the simulation, the potential disagreement with data is considered as the main source of systematic uncertainty. This discrepancy is determined by the following effects:

- A different efficiency in data and simulation for b hadrons of given momentum, due to an imperfect description of tracking and secondary vertex reconstruction. This effect is accounted for by the rescaling factors estimated with the p_T^{rel} method discussed in Section 3.6.3. The uncertainty assigned to them is propagated to the ϵ^{B} corrections, resulting in a global 12% systematic uncertainty on the cross-section values, in all p_T^Z regions.

- A mismodelling of the b-hadron momentum spectrum in the simulation. This effect is physics-process dependent, and is estimated by comparing the average p_T of the subleading reconstructed B candidate in data and MC, in each of the differential cross-section bins, as in Fig. 4.22. The maximal discrepancy Δp_T is approximately 1 GeV, for an average p_T of about 20 GeV. Δp_T is propagated into the corresponding efficiency bias $\Delta\epsilon^{\mathrm{B}}$, using the b-hadron pair identification efficiency curve as a function of the subleading B-candidate transverse momentum shown in Fig 4.22. The resulting uncertainty is about 5%. As this is largely correlated with the uncertainty from the estimation of the IVF

efficiency scale factors, it is not considered as an independent source, and a total 12% uncertainty is assigned to the cross-section values.

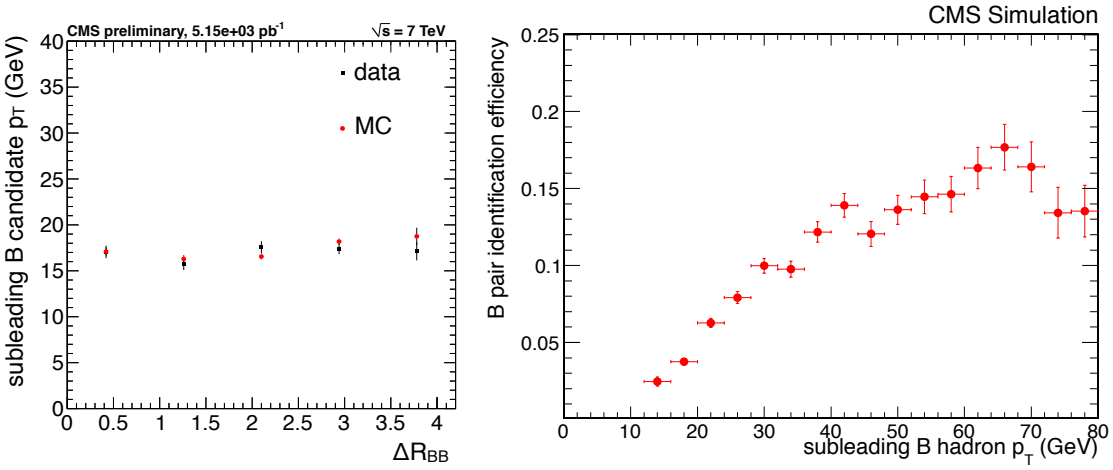

Figure 4.22: *Left*: Average transverse momentum of the subleading B candidate, as a function of ΔR_{BB}. *Right*: Efficiency for tagging a B-hadron pair as a function of the subleading B-hadron p_T, for the inclusive Z sample.

B purity and soft b-hadron contamination

Concerning the $\mathcal{P}$ correction factors, the processes giving the most significant impurity contributions are pp $\rightarrow$ Zb$\bar{b}$X with a misidentified b $\rightarrow$ c decay, and pp $\rightarrow$ Zb$\bar{b}$cX. As b-hadron decays are well known, and reliably reproduced by the simulation, the main systematic uncertainty on the purity correction factors is considered to be related to the latter process. To assess the impact of mismodelling, the $\mathcal{P}$ corrections are estimated varying the fraction of pp $\rightarrow$ Zb$\bar{b}$cX events by $\pm$50%. The IVF algorithm shows robustness against such effect, and the resulting systematic uncertainty on the cross-section estimation is approximately 0.5%.

The soft b-hadron contamination is determined by the poor IVF momentum resolution. As it is evaluated using the simulation, it is affected by a potential mismodelling of the b-hadron kinematics, as well as by an imperfect description of the secondary vertex reconstruction. These sources of uncertainty are already taken into account in the systematic error assigned to the b-hadron pair identification efficiency.

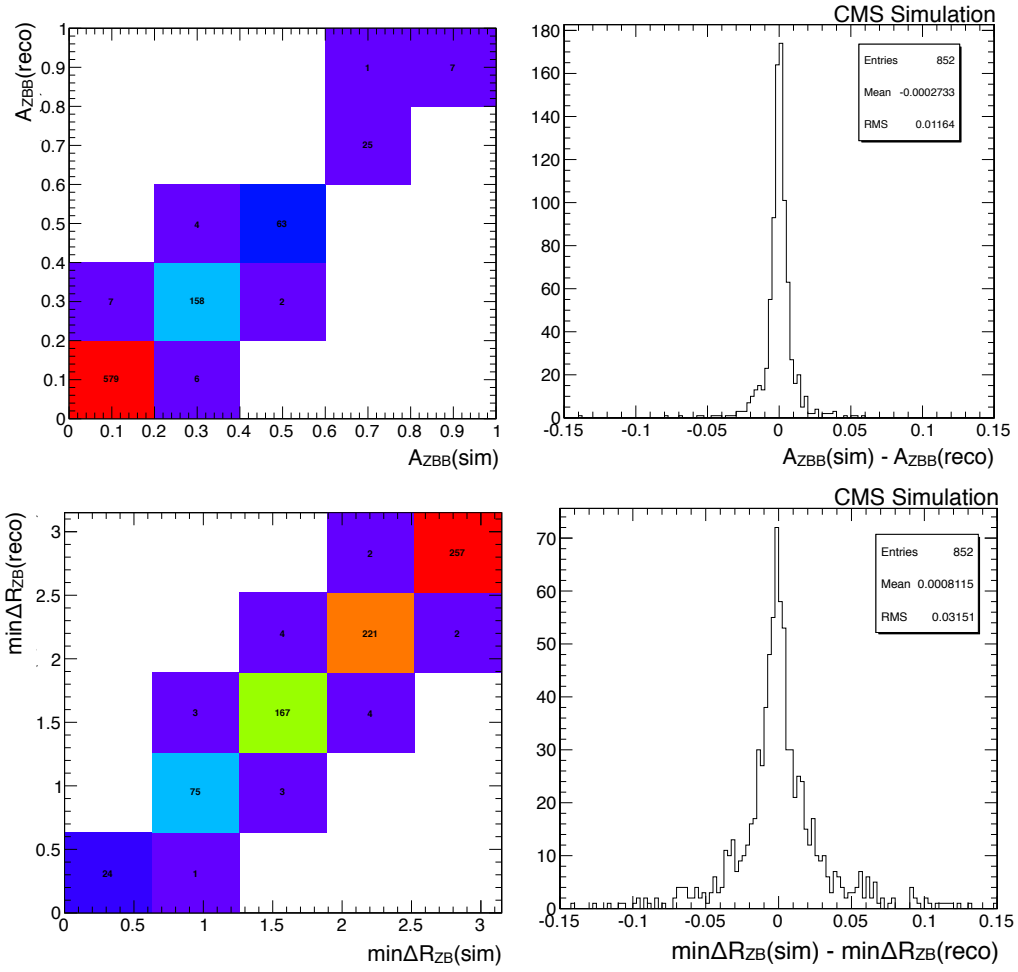

Figure 4.23: Resolution of the angular variables describing the correlations between the Z boson and the b-hadrons: reconstructed $\min\Delta R_{ZB}$ (*top*) and A_{ZBB} (*bottom*) as a function of the true values (*left*), and projection onto the diagonal (*right*). Events with exactly two B candidates and two simulated b hadrons are used. No cut on the Z boson p_T is applied. Similar results are found for the event sample satisfying the $p_T^Z > 50$ GeV requirement.

Additional systematic uncertainties and combination

In addition to the systematic uncertainties described above, three further sources are considered:

- The estimation of the total integrated luminosity ($\mathcal{L}$) collected by the CMS experiment and processed for this analysis. CMS measures the integrated luminosity by counting the average number of hits in the pixel detector in events with a detected proton-proton

collision. The corresponding uncertainty is 2.2% [73];

- The limited size of the available signal MC sample used for the extraction of the b-hadron pair identification efficiency, purity and soft b-hadron contamination corrections. The error is estimated as the square root of the number of entries in each bin of the differential cross-section histograms, and of the size of the entire sample of events passing the selection for the total cross-section measurement. The resulting uncertainty is between 2 and 5% in the differential cross sections, and between 1% and 3% in the total cross-section estimation.

- The migrations of events from one bin of the differential cross-section to the adjacent ones, due to the finite resolution of the correlation angles. For each bin, the fraction of events from other bins is estimated as the ratio between the resolution of the measured variable, and the corresponding bin width. For the ΔR_{BB} and $\Delta\phi_{BB}$ variables that are purely related to the b-hadron reconstruction, the results of a resolution study are shown in Section 3.6, and the systematic uncertainties are between 2% and 3%. For the $\min\Delta R_{ZB}$ and A_{ZBB} variables, describing the correlation between the b-hadron system and the Z boson, the resolution in determining the dilepton momentum direction has to be taken into account as well. The combined IVF and dilepton angular resolution is evaluated using MC simulation. Fig. 4.23 shows the comparison between the reconstructed angles and their true values, for the inclusive Z boson sample. The projection of the two-dimensional histogram onto the diagonal are approximately Gaussian, the width of which provides an estimate of the angular resolution. This is measured to be 0.03 and 0.01 for $\min\Delta R_{ZB}$ and A_{ZBB} respectively, leading to a systematic uncertainty between 2% and 4%. Similar values are found for $p_T^Z > 50$ GeV.

Another potential source of systematic uncertainty, affecting the estimation of dilepton efficiency, b-identification efficiency, purity and soft b-hadron corrections, is the pileup reweighting procedure described in Section 4.2. The correction factors are in fact dependent on pileup, and are evaluated using the MC simulation, which does not exactly reproduce the PV multiplicity distribution observed in data. To estimate the impact of pileup on the final result, the extraction of the hadron-level differential cross sections is independently applied to two event samples, with PV multiplicity above or below 7. The resulting difference is substantially negligible, around 1%. Therefore, no systematic uncertainty from this source is associated to the measurement.

A summary of all systematic uncertainties is given in Table 4.8 and 4.9. These errors are assumed to be uncorrelated, and are therefore combined quadratically with each other, and with the statistical error provided by the BLUE software. The resulting total uncertainty is about 15% for the most populated bins of the differential cross section histograms.

Source	Uncertainty
dilepton channels combination $\mathcal{F}$	$\pm$2%
IVF efficiency scale factors	$\pm$12%
B purity $\mathcal{P}$	$\pm$0.5%
bin-to-bin migrations	from $\pm$1-2% (ΔR_{BB}, minΔR_{ZB}) to $\pm$3-4% ($\Delta\phi_{BB}$, A_{ZBB})
MC statistics	from $\pm$2% to $\pm$3.7%
luminosity	$\pm$2.2%

Table 4.8: Summary of systematic uncertainties assigned to the angular differential cross-section measurements.

Source	Uncertainty
dilepton channels combination $\mathcal{F}$	$\pm$2%
IVF efficiency scale factors	$\pm$12%
B purity $\mathcal{P}$	$\pm$0.5%
MC statistics	from $\pm$1% to $\pm$3.5%
luminosity	$\pm$2.2%

Table 4.9: Summary of systematic uncertainties assigned to the total cross-section estimation.

4.4 Theoretical predictions and uncertainties

The measured cross sections are compared to the tree-level predictions by MadGraph in the four and five-flavour schemes (indicated as MG 4F and MG 5F), and by the four-flavour Alpgen, and to the next-to-leading-order prediction by aMC@NLO. More details about the different MC generators are found in Section 1.3.5. To emulate the effect of higher-order corrections, the MG 5F cross section is rescaled by a k-factor of 1.23, corresponding to the ratio between the next-to-next-to-leading order prediction of the inclusive Z production cross section, and the tree-level cross section estimated by MadGraph. The k-factor for the four-flavour MadGraph (Alpgen) $Zb\bar{b}$ sample is estimated instead by dividing the NLO aMC@NLO cross section ($\sigma = 16.1$ pb for $M_{\ell\ell} > 30$ GeV) by the corresponding tree-level MadGraph (Alpgen) value.

The theoretical uncertainties affecting the MC predictions have to be taken into account when comparing them to the data. The main sources of uncertainties are listed below:

- For the MC implementing the four-flavour scheme, the mass of the b-quark (m_b), set in the generator to 4.7 GeV. It is varied from 4.4 to 5.0 GeV, where $\pm$0.3 GeV is the uncertainty on m_b. This uncertainty is associated to the four-flavour predictions only;

- The choice of the unphysical factorisation and renormalisation scale parameters. The

uncertainty is evaluated by varying their values simultaneously by a factor of two;

- The choice of the unphysical matching scale. Its value, reported in Section 1.3.5, is varied by 15% for the MG 4F generator, and by a factor of two for MG 5F;
- The modelling of the gluon splitting $g \rightarrow b\bar{b}$. To estimate this, the scale at which the strong coupling α_S is evaluated at the splitting vertex is changed from k_T^2 to $m_T^2(b\bar{b}) = m_b^2 + k_T(b\bar{b})$;
- The choice of the PDF. Differential cross sections obtained adopting PDFs from different collaborations are compared. For the four-flavour MadGraph (Alpgen), the MSTW2008LO (CTEQ5L) set is replaced with the MRST2004 (CT10). For the aMC@NLO generators, the uncertainty is estimated by substituting the MSTW2008 set with CT10. For MG 5F the uncertainty is assessed by replacing the CTEQ6L1 set with CT10. The impact on the ΔR_{BB} shape is observed to be negligible. The effect on the cross section normalisation is included in the predicted cross section uncertainty discussed in the next item;
- The uncertainty on the predicted cross section normalisation. For the four-flavour MadGraph and Alpgen it is given by the aMC@NLO uncertainty, obtained varying the factorisation and renormalisation scales simultaneously by a factor of two, and by replacing the MSTW2008 PDF set with CT10. For the five-flavour MadGraph generator the normalisation uncertainty is given by the corresponding NNLO cross section uncertainty [74];
- For the aMC@NLO generator, the uncertainty related to the parton shower. It is assessed from the difference between the predictions obtained with PYTHIA and HERWIG.

A summary of the theoretical uncertainties on the ΔR_{BB} shape is shown in Fig. 4.24. The uncertainty on the cross-section normalisation is not reported in the figure. The predicted ΔR_{BB} trend appears to be heavily affected by the modelling of the gluon splitting, in particular by the choice of the scale at which the coupling α_S is evaluated, which is arbitrary. The adoption of a different scale convention can enhance the collinear b-hadron production by almost a factor of two, and the total cross section by approximately one third. The $g \rightarrow b\bar{b}$ description is therefore the key to improve the goodness of the MC predictions.

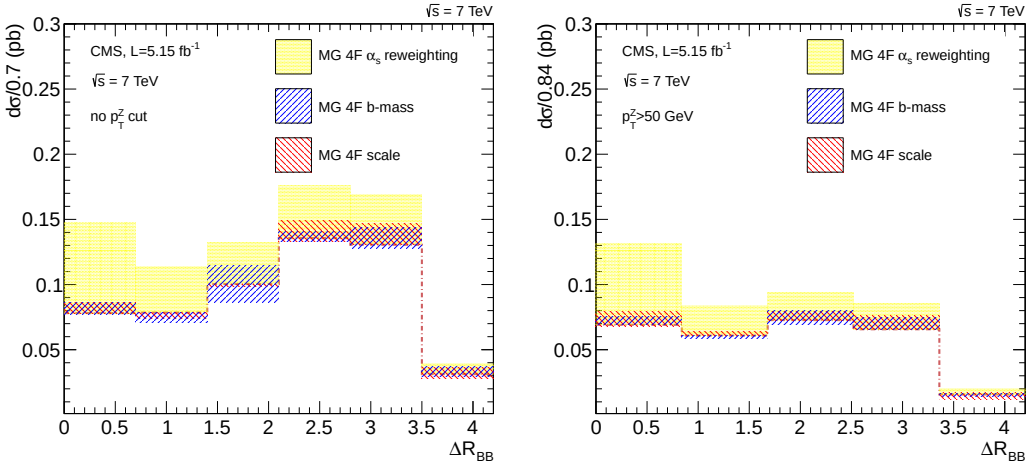

Figure 4.24: Differential $Zb\bar{b}$ cross section as a function of ΔR_{BB}, with no cut on p_T^Z (*left*), and for $p_T^Z > 50$ GeV (*right*), predicted by the four-flavour MadGraph generator, showing the theoretical uncertainties related to the gluon splitting modelling, the b-quark mass and the matching scale.

4.5 Results and interpretation

The total $Zb\bar{b}$ production cross section, and the differential cross sections as a function of the four angular variables ΔR_{BB}, $\Delta\phi_{BB}$, $\min\Delta R_{ZB}$, and A_{ZBB}, are evaluated in the phase space defined by:

- Lepton $p_T^\ell > 20$ GeV and $|\eta^\ell| < 2.4$;
- Dilepton invariant mass $81 < M_{\ell\ell} < 101$ GeV;
- b-hadron $p_T^B > 15$ GeV and $|\eta^B| < 2.0$.

The measured cross sections are compared to the predictions by the MC generators described in Section 1.3.5 and 4.4.

4.5.1 Angular correlations

Results are shown in Fig. 4.25 and 4.26, for the two p_T^Z regions: for the inclusive Z sample, and for $p_T^Z > 50$ GeV. The data, represented as black circles, are compared to the theoretical predictions, shown as triangles and squares.

In both p_T^Z regions the best description of the measured ΔR_{BB} distribution is provided by the four-flavour Alpgen generator. The four and five-flavour MadGraph as well as aMC@NLO

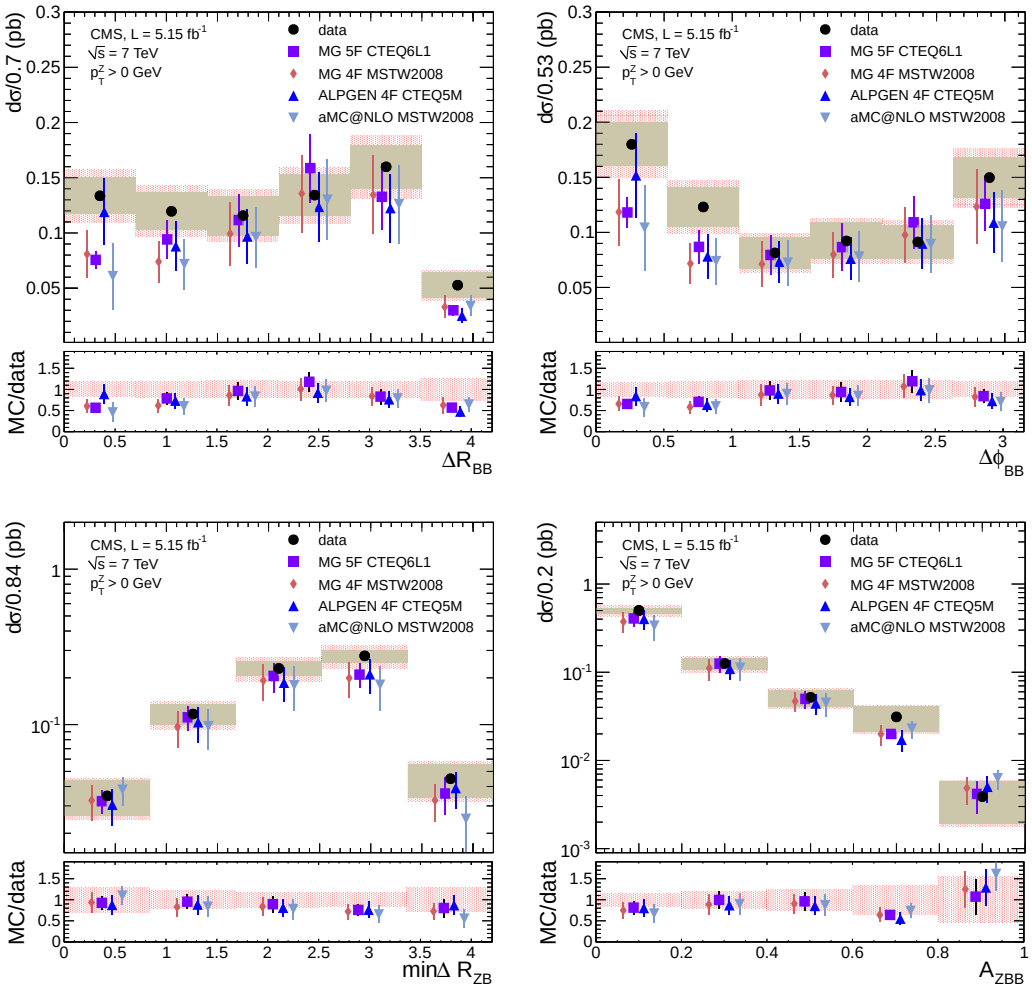

Figure 4.25: Differential Zbb̄ cross section for the inclusive p_T^Z spectrum, as a function of the four angular variables: ΔR_{BB} (*top left*), $\Delta\phi_{BB}$ (*top right*), $\min\Delta R_{ZB}$ (*bottom left*) and A_{ZBB} (*bottom right*). Data are shown as black circles. The red hatched error bands correspond to the quadratic combination of statistical and systematic uncertainty, the brown solid bands show the statistical component only. The MC predictions are represented as squares and triangles.

underestimate the data in the collinear region by 25 to 40%. However, taking into account the theoretical uncertainty related to the scale at which α_S is evaluated in the gluon splitting vertices, also the four-flavour MadGraph prediction is compatible with the data. In the large ΔR_{BB} region, the MC well reproduces the measurement.

Similar conclusions can be drawn about $\Delta\phi_{BB}$. As for ΔR_{BB}, the best agreement with data is achieved by Alpgen. Nevertheless, all the predictions are between one and three sigmas below

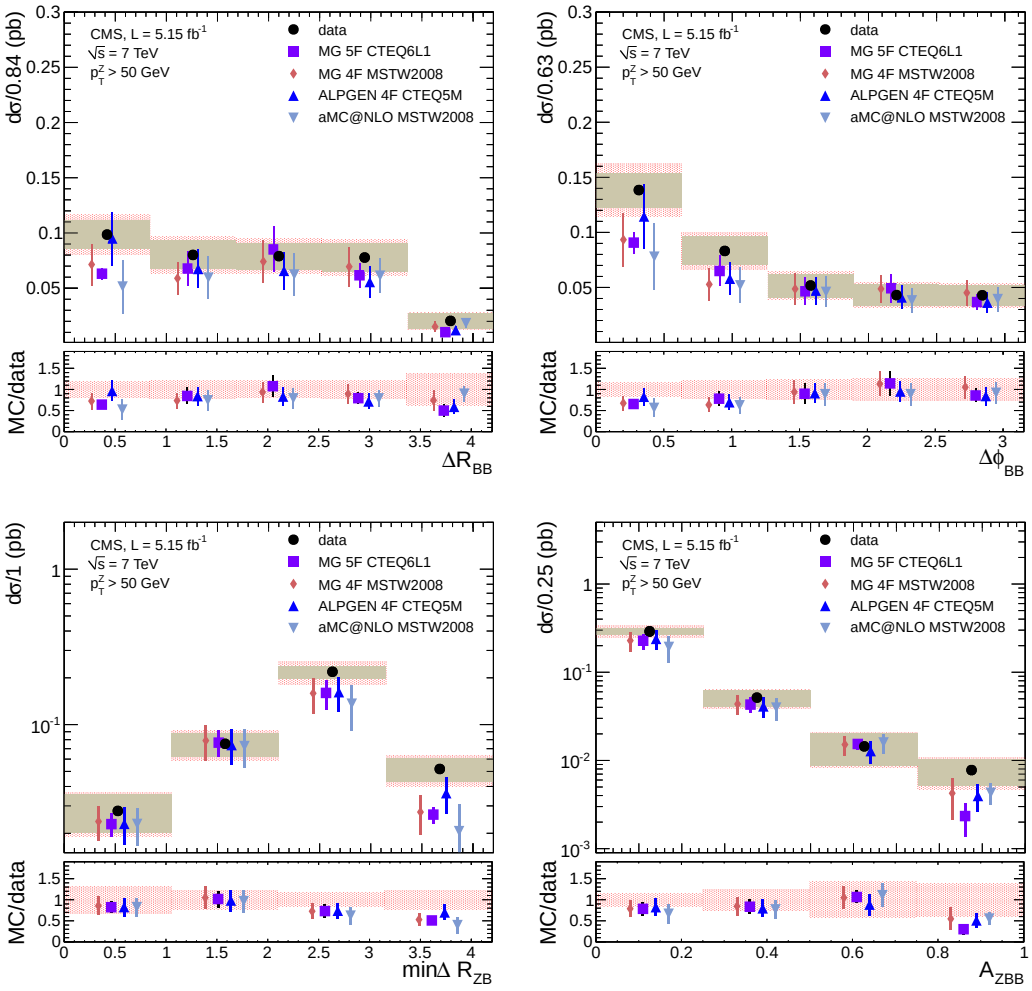

Figure 4.26: Differential Zb$\bar{\text{b}}$ cross section for $p_T^Z > 50$ GeV, as a function of the four angular variables: ΔR_{BB} (*top left*), $\Delta\phi_{BB}$ (*top right*), minΔR_{ZB} (*bottom left*) and A_{ZBB} (*bottom right*). Data are shown as black circles. The red hatched error bands correspond to the quadratic combination of statistical and systematic uncertainty, the brown solid bands show the statistical component only. The MC predictions are represented as squares and triangles.

the data in the region corresponding to the collinear b-hadron production. Also in this case, the measurement is in agreement with the four-flavour MadGraph generator within the α_S scale uncertainty. The discrepancy is less pronounced in the $p_T^Z > 50$ GeV region.

The correlation between the Z boson and the b hadrons is in general fairly reproduced by the MC generators: all predictions are in agreement with the minΔR_{ZB} and A_{ZBB} measurements. Some tension, approximately one sigma, is observed however for min$\Delta R_{ZB} > 2.5$ and $A_{ZBB} <$

0.2, in both p_T^Z regions.

The four differential distributions seem to coherently suggest a mismodelling of the processes involving a gluon splitting vertex ($qq \rightarrow Zb\bar{b}X$ and $qg \rightarrow Zb\bar{b}X$), discussed in the introduction to this chapter. This hypothesis is confirmed by the agreement achieved when modifying the α_S scale at the gluon splitting vertices.

4.5.2 Total cross section

The measurement of the angular correlations is completed by the estimation of the total $Zb\bar{b}$ production cross section in the four p_T^Z regions: for the inclusive Z sample, and for $p_T^Z > 40$, 80 and 120 GeV. The result is shown in Fig. 4.27. The data are represented as black circles, sur-

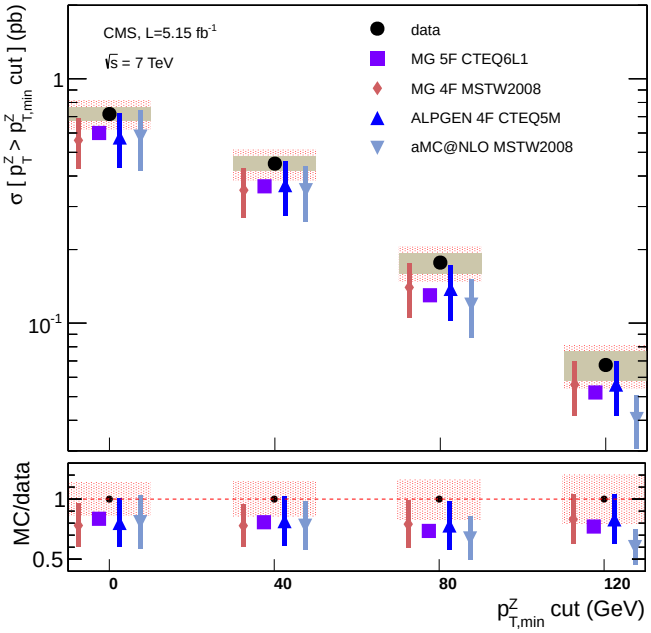

Figure 4.27: Total $Zb\bar{b}$ production cross section as a function of the p_T^Z cut. The data is shown as black circles. The red hatched error band represents the quadratic combination of statistical and systematic uncertainties, while the solid brown band corresponds to the statistical uncertainty only. The MC predictions are shown as squares and triangles. The bottom frame reports the ratio between the MC prediction and the data.

rounded by error bands representing the statistical and systematic uncertainties. All the predictions are lower than the data: the tree-level MC generators underestimate the measurement

by approximately 15%, the NLO aMC@NLO generator by 50%. The trend as a function of the p_T^Z cut is reproduced by the four-flavour MadGraph and Alpgen, while a discrepancy at high p_T^Z is observed in the comparison with MadGraph in the five-flavour scheme and aMC@NLO.

CONCLUSIONS

This thesis presented the measurement of the angular correlations and the total cross section of the associated production pp $\rightarrow$ Zb$\bar{\text{b}}$X, based on a proton-proton collision data sample corresponding to an integrated luminosity of 5.15 fb^{-1} at a center-of-mass energy $\sqrt{s} = 7$ TeV, collected by the CMS experiment at the LHC.

The associated production of vector bosons and heavy quarks is one of the major backgrounds to the Standard Model Higgs boson production. However, it is not well understood, and large theoretical uncertainties especially affect the modelling of the production of collinear b-quark pairs. Different calculation schemes are adopted by Monte Carlo generators for the simulation of the Zb$\bar{\text{b}}$ final state: the four-flavour scheme, assuming a massive b quark, and the five-flavour approximation, considering the b quark as massless. The analysis of the angular correlations allows to identify the approximation that provides the best description of the data.

The differential production cross section is measured as a function of four angles parametrizing the correlations between the b hadrons and the Z. The reference variable is the three dimensional angular separation between the b-hadron flight directions, ΔR_{BB}, which is sensitive to the Zb$\bar{\text{b}}$ production mode. It allows to probe the modelling of the subprocesses qq $\rightarrow$ Zb$\bar{\text{b}}$X and qg $\rightarrow$ Zb$\bar{\text{b}}$X, associated to Feynman diagrams with a gluon splitting vertex g $\rightarrow$ b$\bar{\text{b}}$. The ΔR_{BB} information is complemented by the transverse angular separation between the two b hadrons, $\Delta\phi_{BB}$. The correlations between the Z and the b hadrons are parametrized by two variables: the three-dimensional separation between the Z momentum and the flight direction of the closest b hadron minΔR_{ZB}, and the asymmetry A_{ZBB}, defined as $A_{ZBB} = \frac{\max\Delta R_{ZB} - \min\Delta R_{ZB}}{\max\Delta R_{ZB} + \min\Delta R_{ZB}}$. These variables constitute a test for the modelling of the gluon splitting component, and are sensitive to the emission of additional QCD radiation.

The cross sections are measured in various kinematic regions, defined by cuts on the Z boson transverse momentum: $p_T^Z > 0$, 40, 80, 120 GeV for the total cross section, and $p_T^Z > 0$, 50 GeV for the differential cross sections.

The Zb$\bar{\text{b}}$ final state is selected by requiring the presence of two isolated high p_T muons or electrons with invariant mass in a narrow interval centred in the Z boson mass, and exactly two b-hadron candidates. The b hadrons are identified using the inclusive vertex finder method

documented in Section 3.6. The IVF algorithm, independent from jet reconstruction, extends for the first time the sensitivity to the collinear b-hadron pair production, that was experimentally inaccessible so far.

A fraction of the selected event sample is expected to be from $t\bar{t}$ production. The signal yields are extracted with an extended maximum likelihood fit of the dilepton invariant mass distribution. The cross sections are obtained by corrected the detector-level signal yields and distributions to the hadron level. The corrections account for the dilepton and b-hadron pair identification and selection efficiency and acceptance, the contamination from mistagged charmed hadron decays and from b hadrons with p_T lower than the acceptance threshold. Due to the excellent angular resolution provided by the IVF algorithm, the bin-to-bin event migrations are negligible, and the unfolding can be applied independently in each bin.

The statistical uncertainty is approximately 10% in the most populated bins of the differential cross sections. The dominant systematic uncertainty is related to the measurement of the b-hadron identification efficiency with data, and is about 12%.

The measured cross sections are compared to the leading-order predictions by ALPGEN in the four-flavour scheme, and by MADGRAPH in the four and five-flavour schemes, and to the NLO prediction by aMC@NLO. The generator providing the best description of data is ALPGEN. Localised excesses of data with respect to the other predictions are found in the regions of small separation between the b hadrons ($\Delta R_{BB} < 1.2$ and $\Delta\phi_{BB} < 1.0$), and for $\min\Delta R_{ZB} > 2.5$ and $A_{ZBB} < 0.2$. The observation is consistent with the hypothesis of a mismodelling of the contribution from the subprocesses associated to Feynman diagrams with a gluon splitting vertex $g \to b\bar{b}$. For the generators implementing the four-flavour scheme, this is confirmed by a study of the theoretical uncertainties. Different choices of the scale at which the strong coupling α_S is evaluated at the gluon splitting vertices lead to significance variations in the cross sections normalisation and in the angular trends, allowing to achieve good agreement with the data.

Besides the disagreement observed in the differential cross sections, a 15% discrepancy is found in the total cross section normalisation, with the data being above than all predictions. The trend as a function of the p_T^Z cut is on the other hand well reproduced by the generators.

The angular analysis presented in this thesis sheds light on the $pp \to Zb\bar{b}X$ production mechanisms, probing for the first time the collinear b-hadron pair production. Future measurements of the differential cross section with the larger data sample collected by the CMS experiment in 2012, at a center-of-mass energy $\sqrt{s} = 8$ TeV, may allow to state with better precision which Monte Carlo generator and which calculation scheme gives the best description of the data.

Another process of great interest at the LHC is the $pp \to W^{\pm}b\bar{b}X$ associated production, which, together with the $Zb\bar{b}$ final state, is background to the SM Higgs boson search. As for the $Zb\bar{b}$ final state, the theoretical predictions of $W^{\pm}b\bar{b}$ are affected by large uncertainties on the collinear b-quark pair production, related to the gluon splitting mechanism. A measurement

of the angular correlations in the $W^{\pm}b\bar{b}$ at CMS, exploiting the same inclusive vertex finder tool, could therefore confirm the observations discussed in this work.

APPENDIX

A

ADDITIONAL PLOTS AND TABLES

This appendix contains additional plots and tables documenting the calculation of the total and differential cross sections. Section A.1 shows the extended maximum likelihood fits for the extraction of the $Zb\bar{b}$ signal yields. The dilepton and b-hadron reconstruction efficiency and purity corrections are documented in Section A.2.

A.1 Estimation of the signal yields

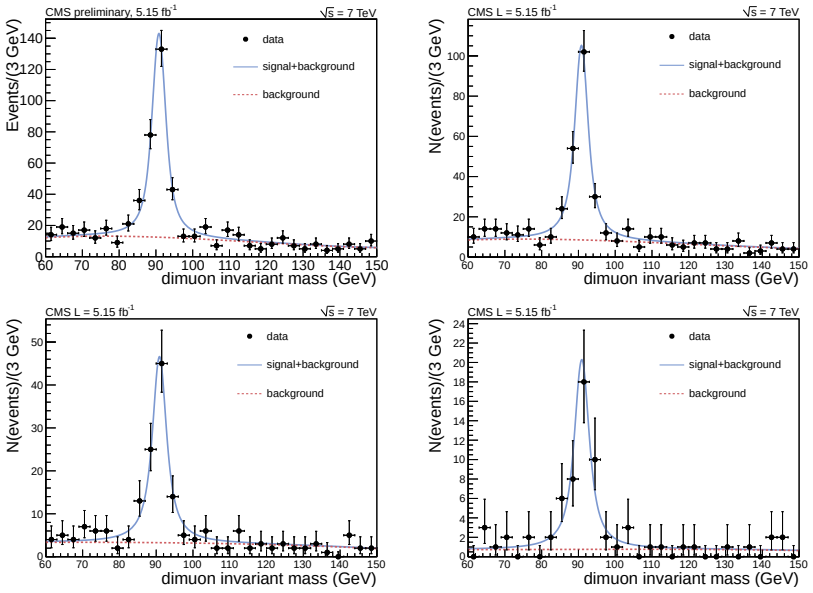

Figure A.1: Extended maximum likelihood fit of the dimuon invariant mass distribution, with no cut on p_T^Z (*top left*), and for $p_T^Z > 40$ GeV (*top right*), 80 GeV (*bottom left*) and 120 GeV (*bottom right*).

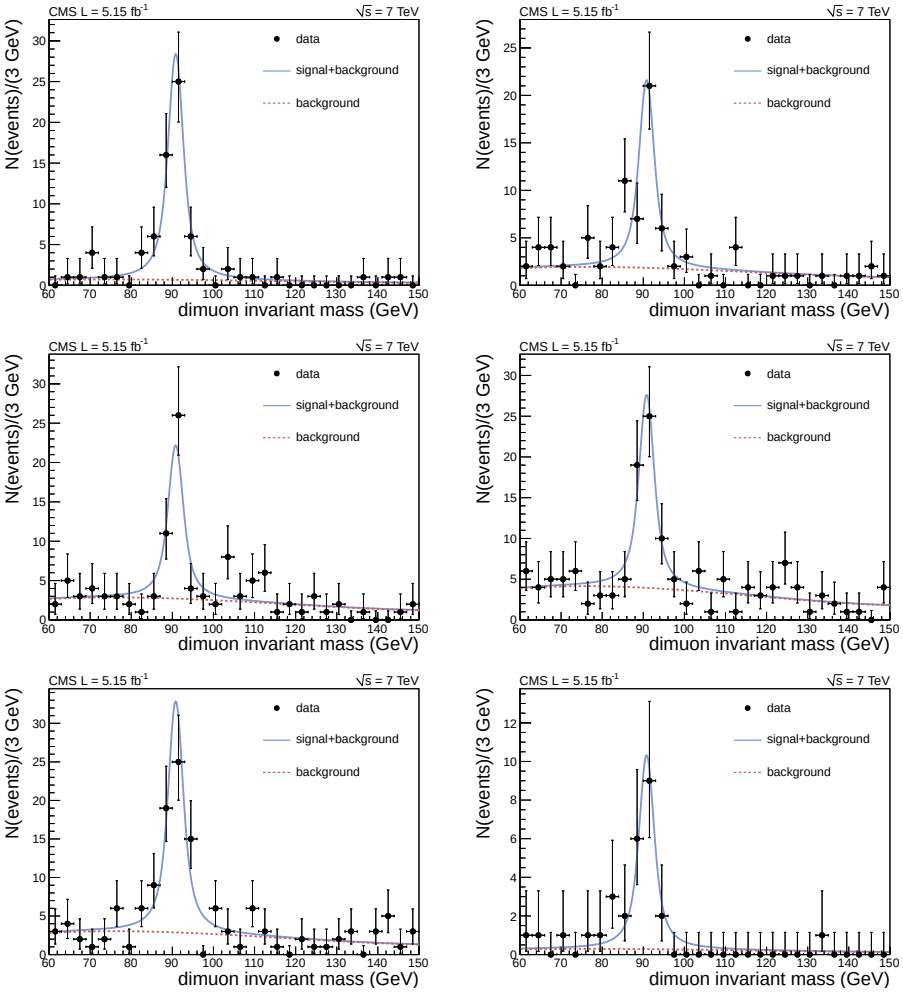

Figure A.2: Extended maximum likelihood fit of the dimuon invariant mass distribution for the signal extraction, in all bins of the differential cross section as a function of ΔR_{BB}, with no cut on p_T^Z. From top left to bottom right: $\Delta R_{BB} < 0.7$, $0.7 < \Delta R_{BB} < 1.4$, $1.4 < \Delta R_{BB} < 2.1$, $2.1 < \Delta R_{BB} < 2.8$, $2.8 < \Delta R_{BB} < 3.5$, $3.5 < \Delta R_{BB} < 4.2$.

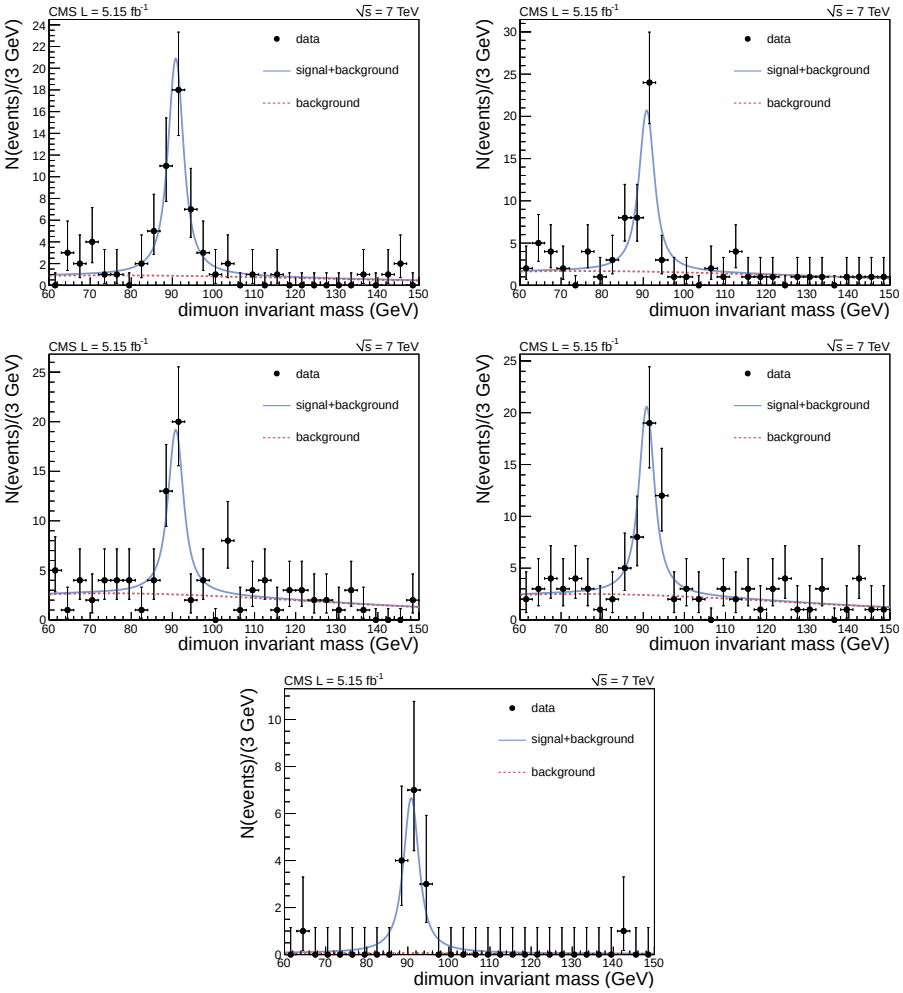

Figure A.3: Extended maximum likelihood fit of the dimuon invariant mass distribution for the signal extraction, in all bins of the differential cross section as a function of ΔR_{BB}, for $p_T^Z > 50$ GeV. From top left to bottom right: $\Delta R_{BB} < 0.84$, $0.84 < \Delta R_{BB} < 1.68$, $1.68 < \Delta R_{BB} < 2.52$, $2.52 < \Delta R_{BB} < 3.36$, $3.36 < \Delta R_{BB} < 4.2$.

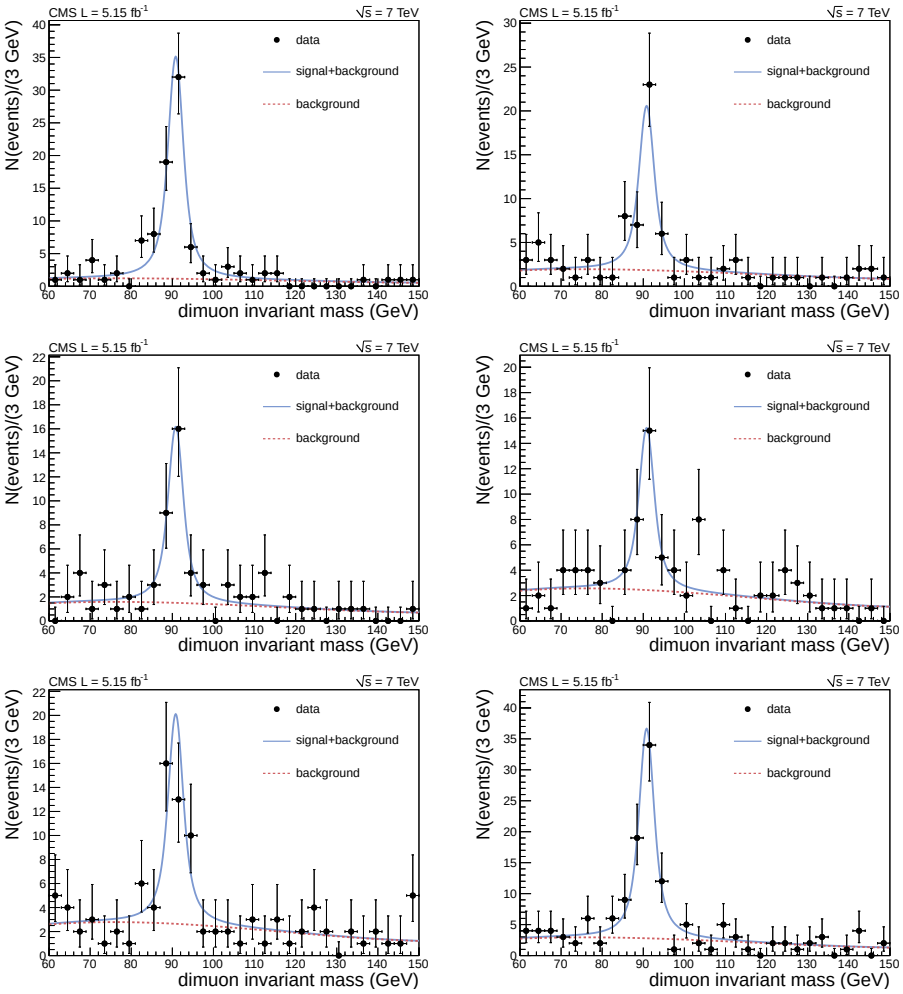

Figure A.4: Extended maximum likelihood fit of the dimuon invariant mass distribution for the signal extraction, in all bins of the differential cross section as a function of $\Delta\phi_{BB}$, with no cut on p_T^Z . From top left to bottom right: $\Delta\phi_{BB} < 0.525$, $0.525 < \Delta\phi_{BB} < 1.05$, $1.05 < \Delta\phi_{BB} < 1.575$, $1.575 < \Delta\phi_{BB} < 2.1$, $2.1 < \Delta\phi_{BB} < 2.625$, $2.625 < \Delta\phi_{BB} < 3.15$.

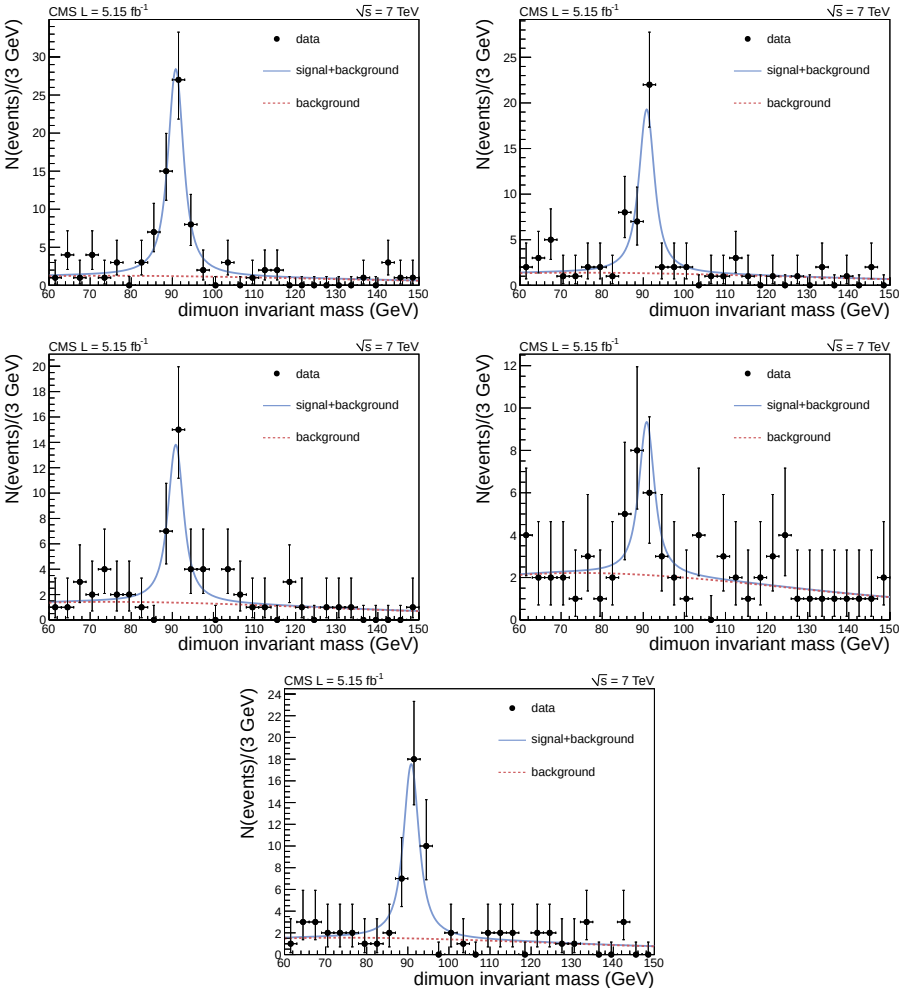

Figure A.5: Extended maximum likelihood fit of the dimuon invariant mass distribution for the signal extraction, in all bins of the differential cross section as a function of $\Delta\phi_{BB}$, for $p_T^Z > 50$ GeV. From top left to bottom right: $\Delta\phi_{BB} < 0.63$, $0.63 < \Delta\phi_{BB} < 1.26$, $1.26 < \Delta\phi_{BB} < 1.89$, $1.89 < \Delta\phi_{BB} < 2.52$, $2.52 < \Delta\phi_{BB} < 3.15$.

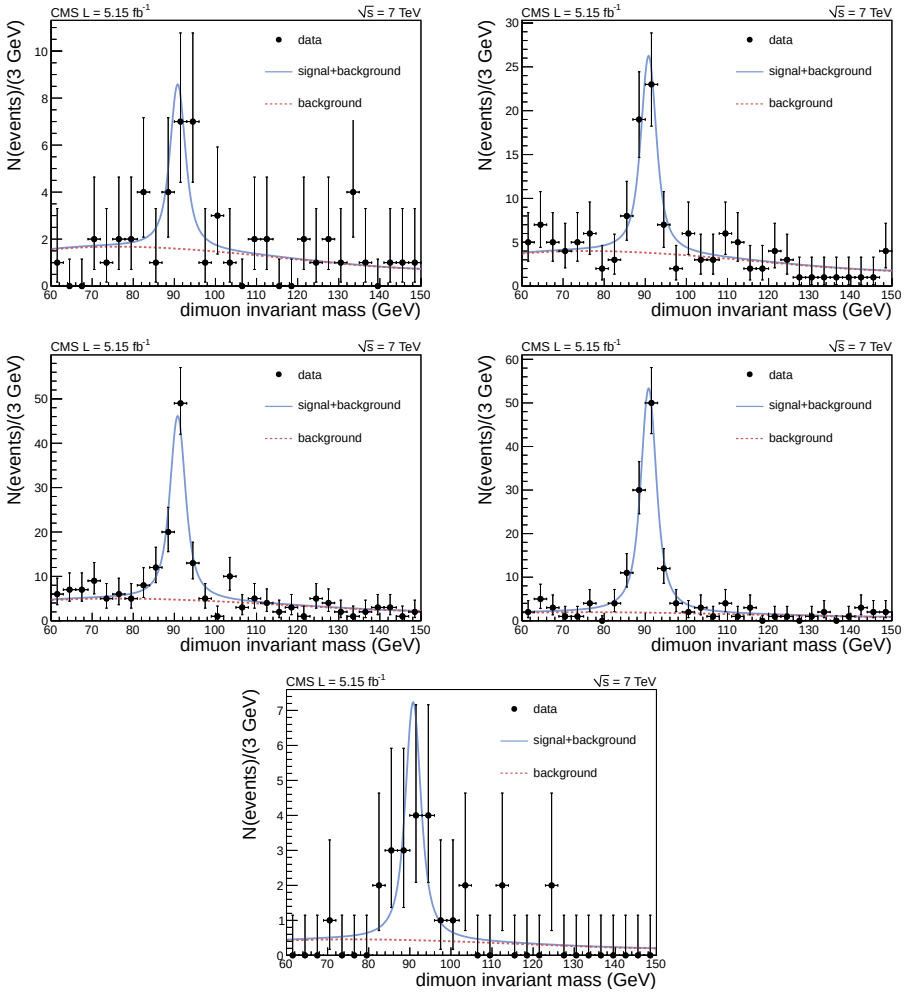

Figure A.6: Extended maximum likelihood fit of the dimuon invariant mass distribution for the signal extraction, in all bins of the differential cross section as a function of $\min\Delta R_{ZB}$, with no cut on p_T^Z. From top left to bottom right: $\min\Delta R_{ZB} < 0.84$, $0.84 < \min\Delta R_{ZB} < 1.68$, $1.68 < \min\Delta R_{ZB} < 2.52$, $2.52 < \min\Delta R_{ZB} < 3.36$, $3.36 < \min\Delta R_{ZB} < 4.2$.

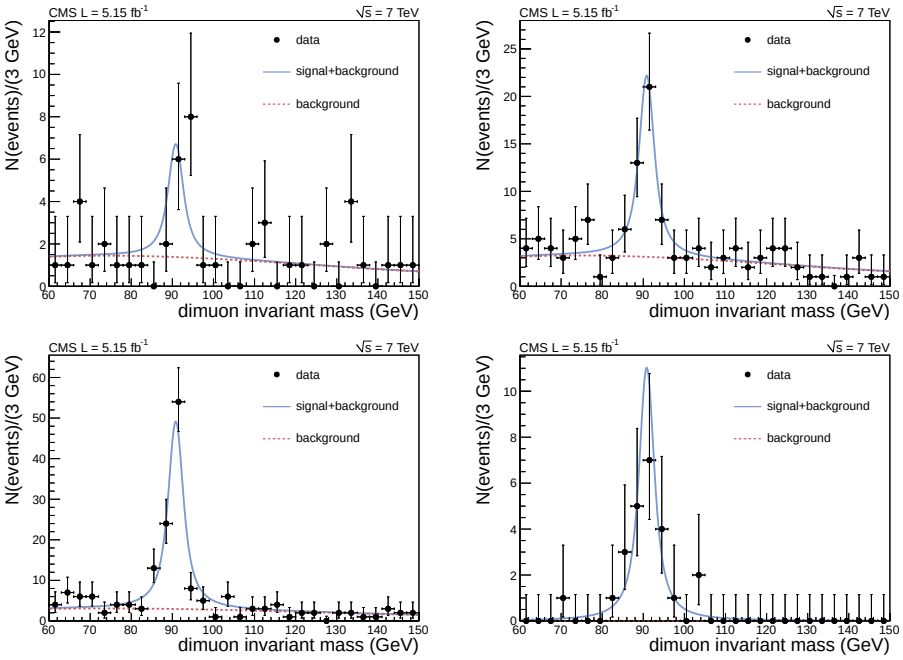

Figure A.7: Extended maximum likelihood fit of the dimuon invariant mass distribution for the signal extraction, in all bins of the differential cross section as a function of $\min\Delta R_{ZB}$, for $p_T^Z > 50$ GeV. From top left to bottom right: $\min\Delta R_{ZB} < 1.05$, $1.05 < \min\Delta R_{ZB} < 2.1$, $2.1 < \min\Delta R_{ZB} < 3.15$, $3.15 < \min\Delta R_{ZB} < 4.2$.

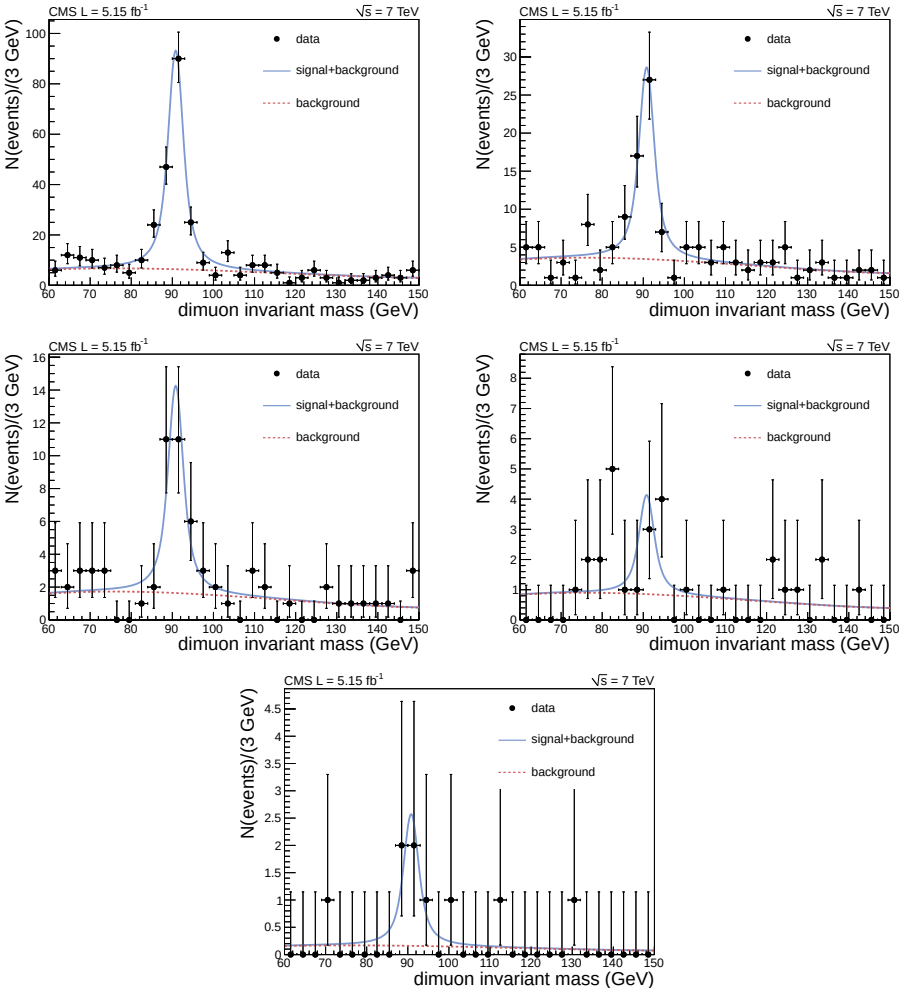

Figure A.8: Extended maximum likelihood fit of the dimuon invariant mass distribution for the signal extraction, in all bins of the differential cross section as a function of A_{ZBB}, with no cut on p_T^Z. From top left to bottom right: $A_{ZBB} < 0.2$, $0.2 < A_{ZBB} < 0.4$, $0.4 < A_{ZBB} < 0.6$, $0.6 < A_{ZBB} < 0.8$, $0.8 < A_{ZBB} < 1.0$.

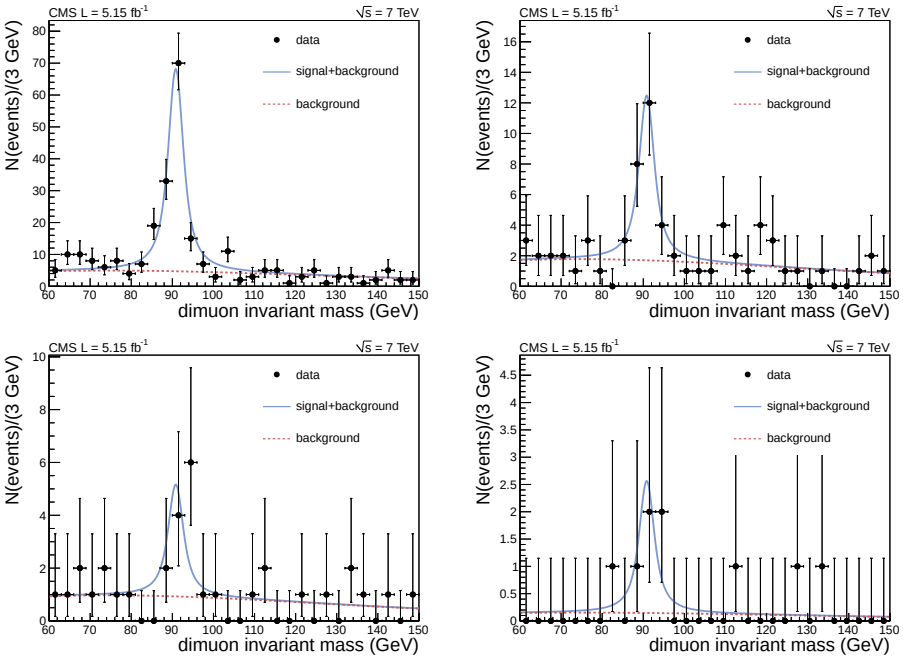

Figure A.9: Extended maximum likelihood fit of the dimuon invariant mass distribution for the signal extraction, in all bins of the differential cross section as a function of A_{ZBB}, for $p_T^Z > 50$ GeV. From top left to bottom right: $A_{ZBB} < 0.25$, $0.25 < A_{ZBB} < 0.5$, $0.5 < A_{ZBB} < 0.75$, $0.75 < A_{ZBB} < 1.0$.

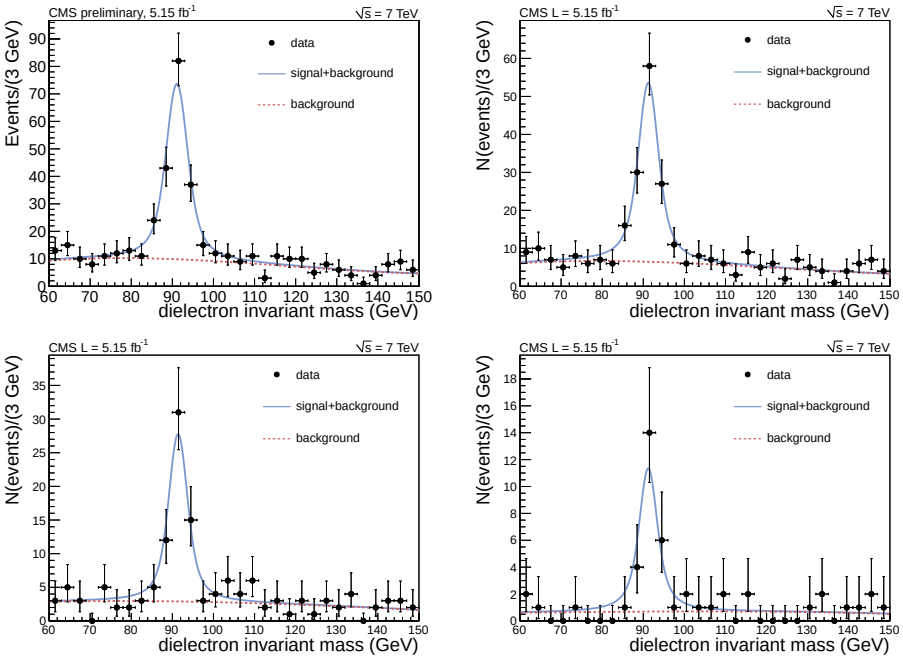

Figure A.10: Extended maximum likelihood fit of the dielectron invariant mass distribution for the signal extraction, with no cut on p_T^Z (*top left*), and for $p_T^Z > 40$ GeV (*top right*), 80 GeV (*bottom left*) and 120 GeV (*bottom right*).

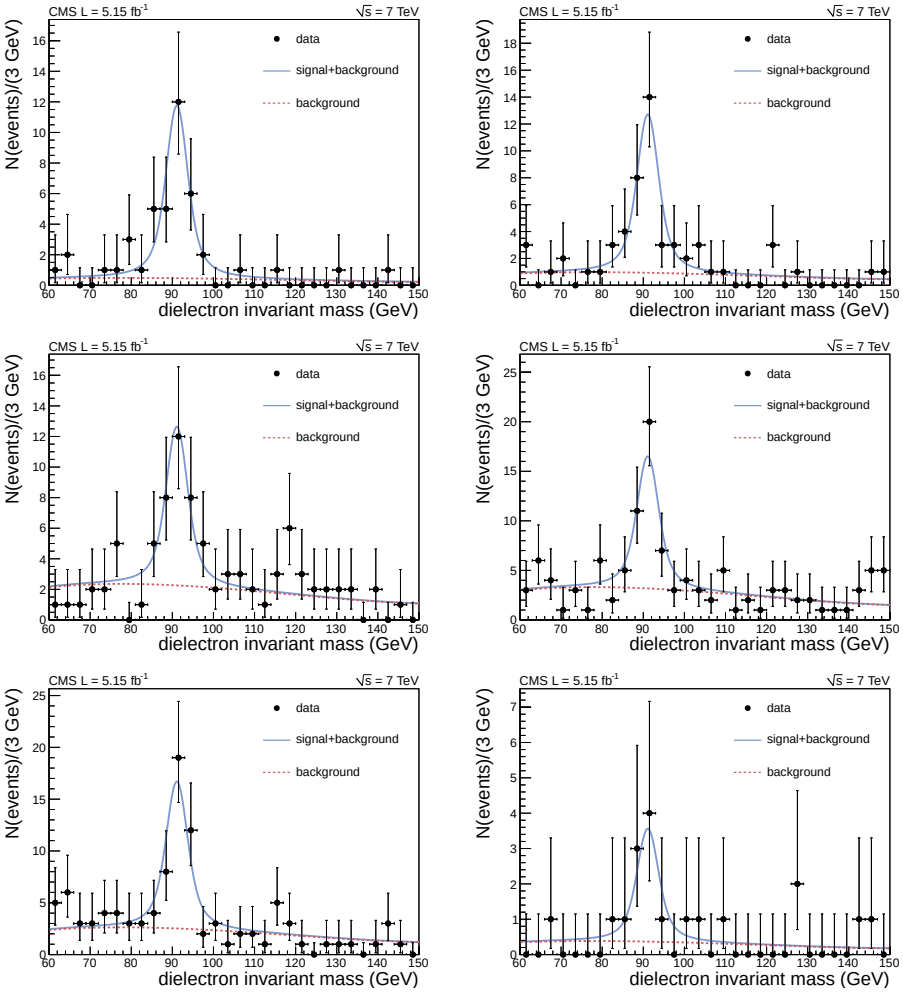

Figure A.11: Extended maximum likelihood fit of the dielectron invariant mass distribution for the signal extraction, in all bins of the differential cross section as a function of ΔR_{BB}, with no cut on p_T^Z. From top left to bottom right: $\Delta R_{BB} < 0.7$, $0.7 < \Delta R_{BB} < 1.4$, $1.4 < \Delta R_{BB} < 2.1$, $2.1 < \Delta R_{BB} < 2.8$, $2.8 < \Delta R_{BB} < 3.5$, $3.5 < \Delta R_{BB} < 4.2$.

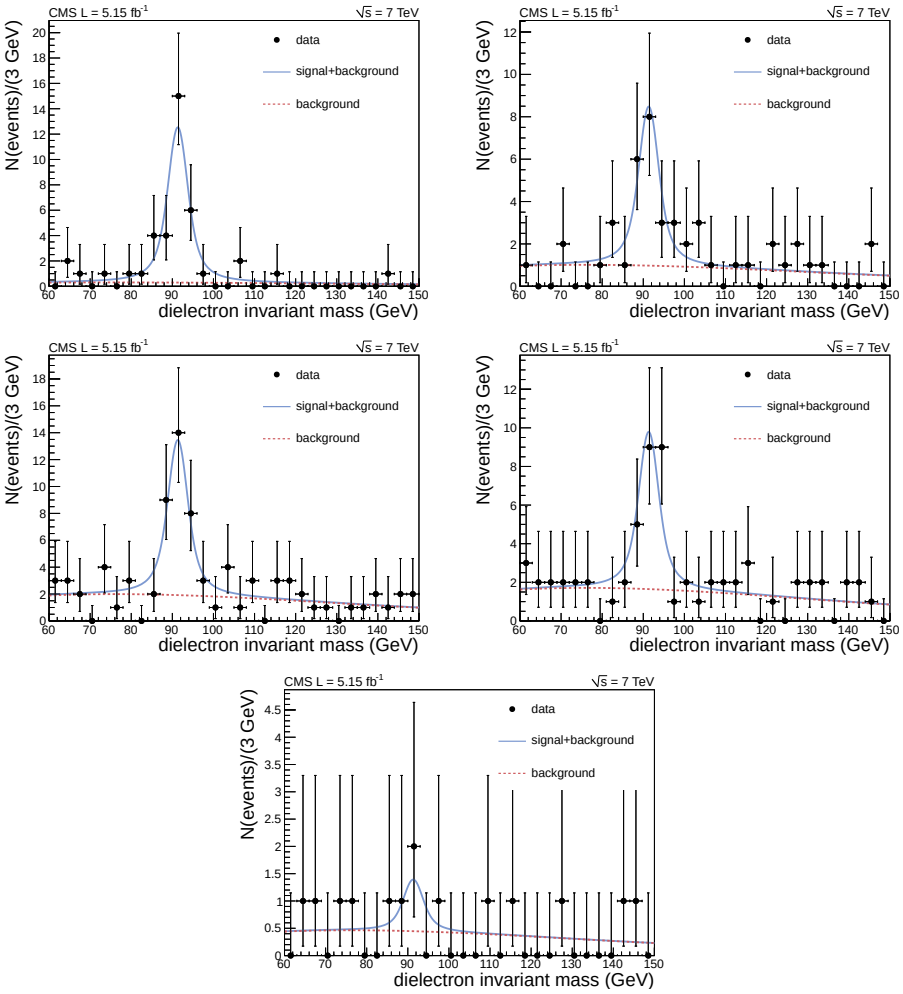

Figure A.12: Extended maximum likelihood fit of the dielectron invariant mass distribution for the signal extraction, in all bins of the differential cross section as a function of ΔR_{BB}, for $p_T^Z > 50$ GeV. From top left to bottom right: $\Delta R_{BB} < 0.84$, $0.84 < \Delta R_{BB} < 1.68$, $1.68 < \Delta R_{BB} < 2.52$, $2.52 < \Delta R_{BB} < 3.36$, $3.36 < \Delta R_{BB} < 4.2$.

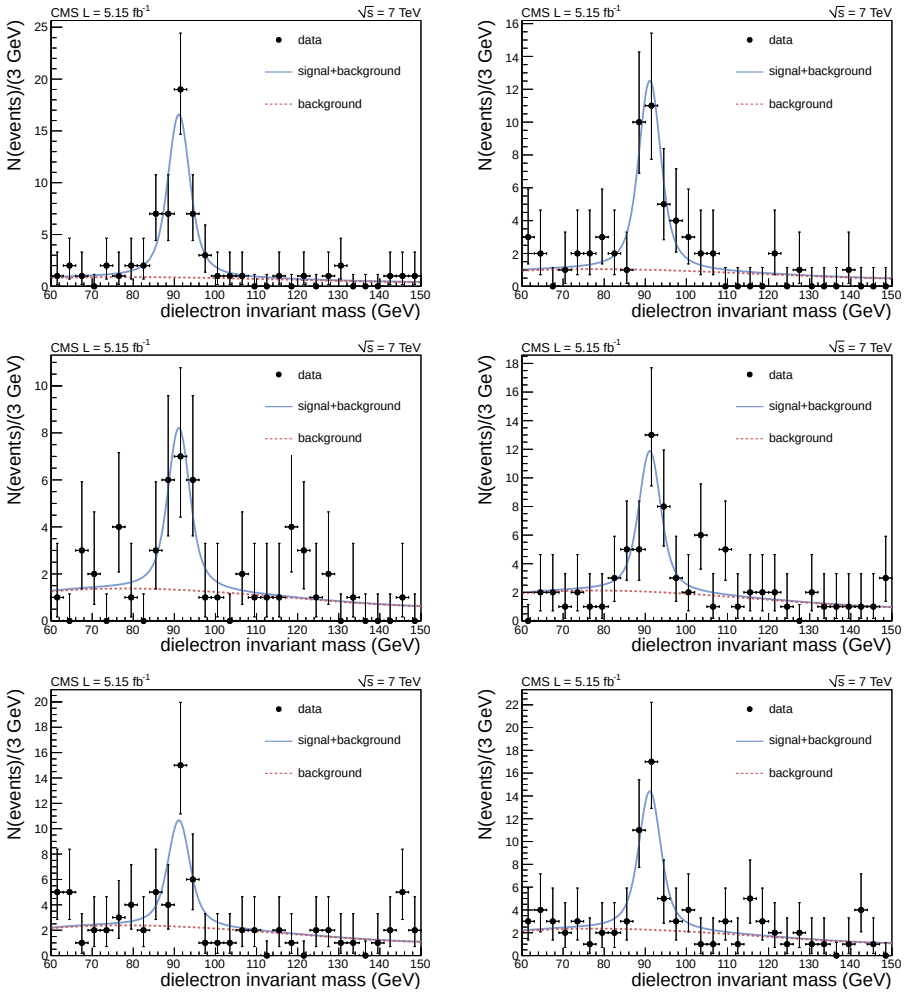

Figure A.13: Extended maximum likelihood fit of the dielectron invariant mass distribution for the signal extraction, in all bins of the differential cross section as a function of $\Delta\phi_{BB}$, with no cut on p_T^Z. From top left to bottom right: $\Delta\phi_{BB} < 0.525$, $0.525 < \Delta\phi_{BB} < 1.05$, $1.05 < \Delta\phi_{BB} < 1.575$, $1.575 < \Delta\phi_{BB} < 2.1$, $2.1 < \Delta\phi_{BB} < 2.625$, $2.625 < \Delta\phi_{BB} < 3.15$.

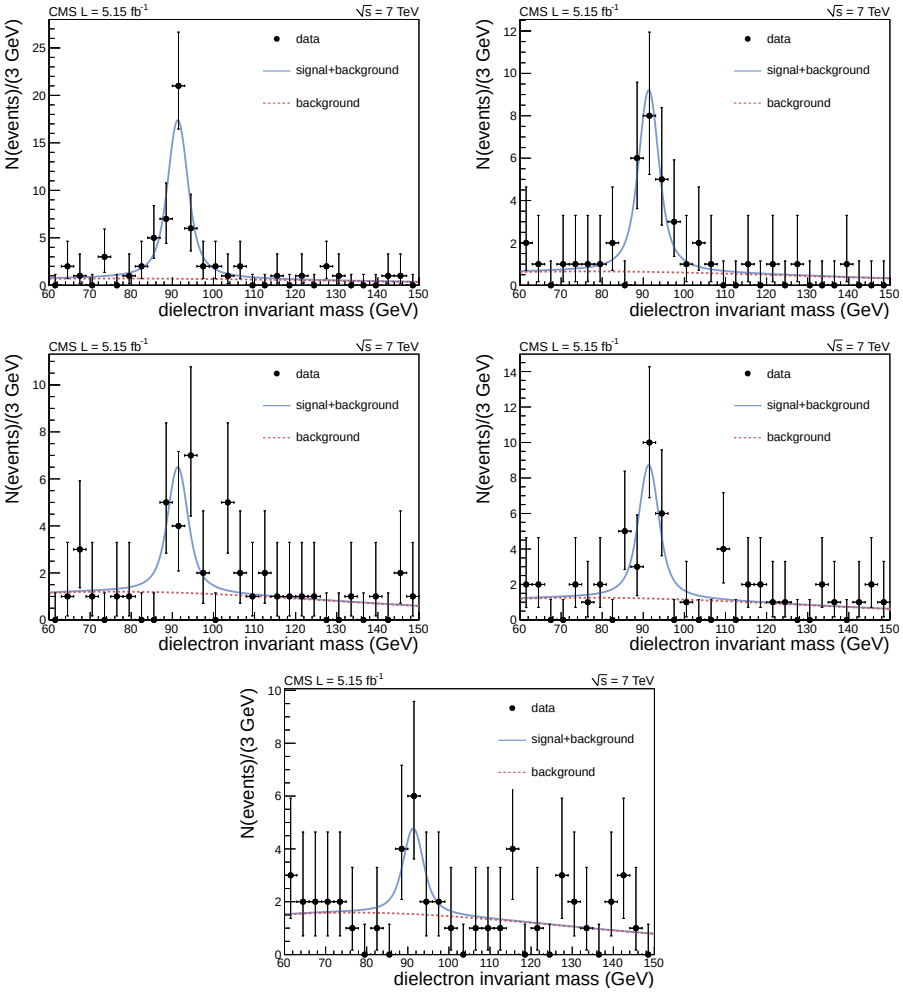

Figure A.14: Extended maximum likelihood fit of the dielectron invariant mass distribution for the signal extraction, in all bins of the differential cross section as a function of $\Delta\phi_{BB}$, for $p_T^Z > 50$ GeV. From top left to bottom right: $\Delta\phi_{BB} < 0.63$, $0.63 < \Delta\phi_{BB} < 1.26$, $1.26 < \Delta\phi_{BB} < 1.89$, $1.89 < \Delta\phi_{BB} < 2.52$, $2.52 < \Delta\phi_{BB} < 3.15$.

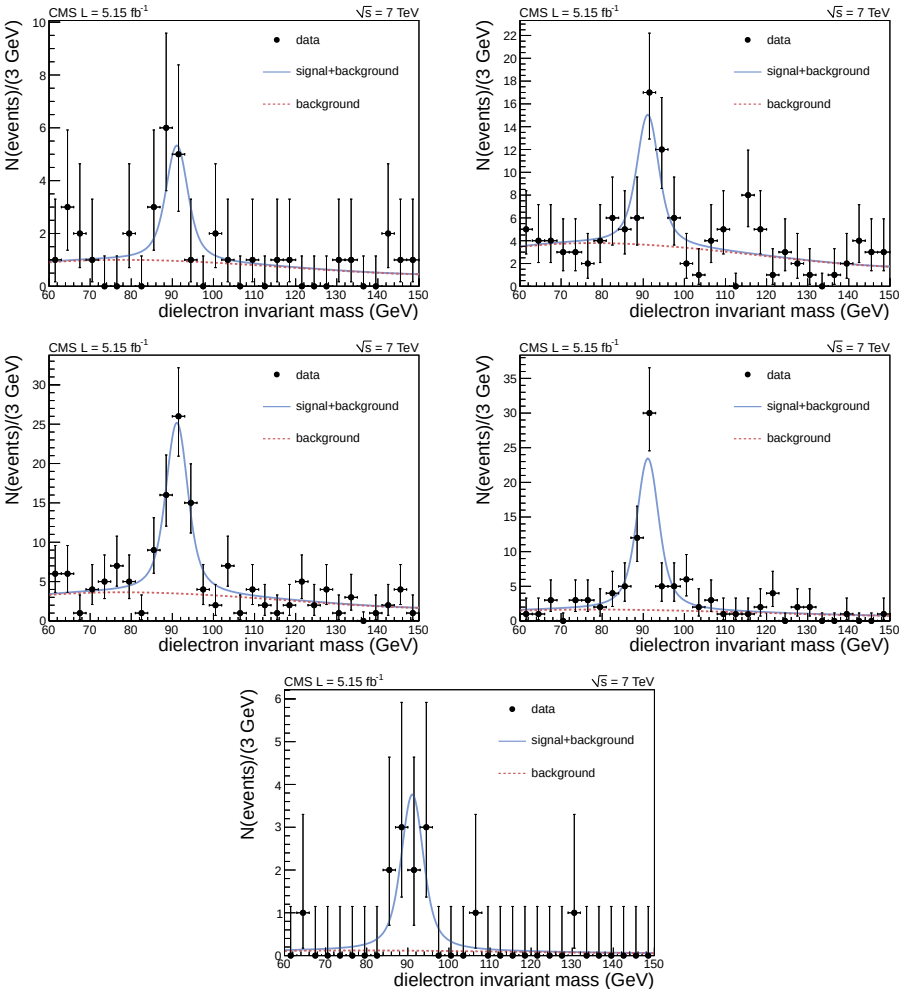

Figure A.15: Extended maximum likelihood fit of the dielectron invariant mass distribution for the signal extraction, in all bins of the differential cross section as a function of $\min\Delta R_{ZB}$, with no cut on p_T^Z. From top left to bottom right: $\min\Delta R_{ZB} < 0.84$, $0.84 < \min\Delta R_{ZB} < 1.68$, $1.68 < \min\Delta R_{ZB} < 2.52$, $2.52 < \min\Delta R_{ZB} < 3.36$, $3.36 < \min\Delta R_{ZB} < 4.2$.

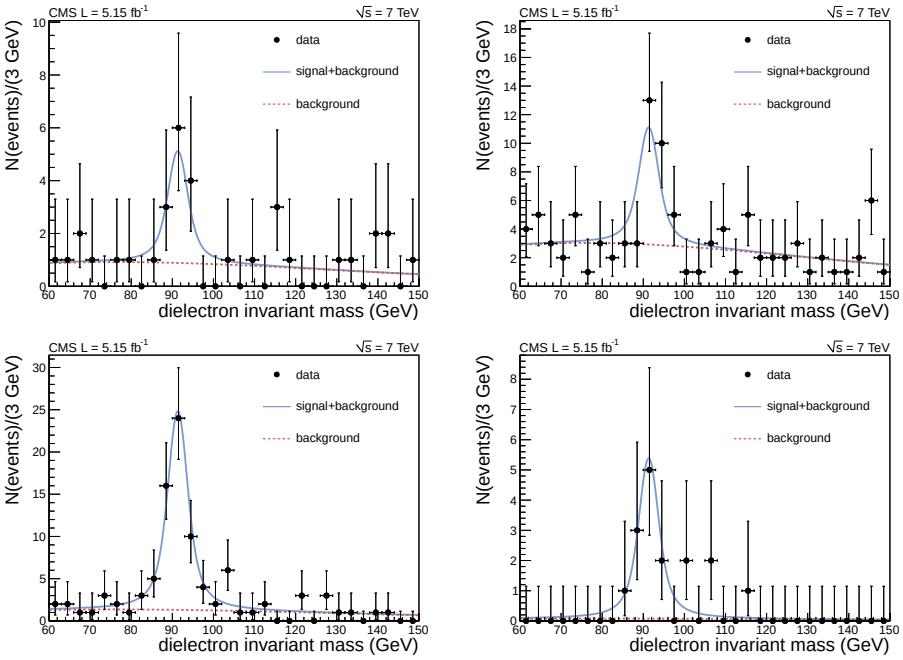

Figure A.16: Extended maximum likelihood fit of the dielectron invariant mass distribution for the signal extraction, in all bins of the differential cross section as a function of $\min\Delta R_{ZB}$, for $p_T^Z > 50$ GeV. From top left to bottom right: $\min\Delta R_{ZB} < 1.05$, $1.05 < \min\Delta R_{ZB} < 2.1$, $2.1 < \min\Delta R_{ZB} < 3.15$, $3.15 < \min\Delta R_{ZB} < 4.2$.

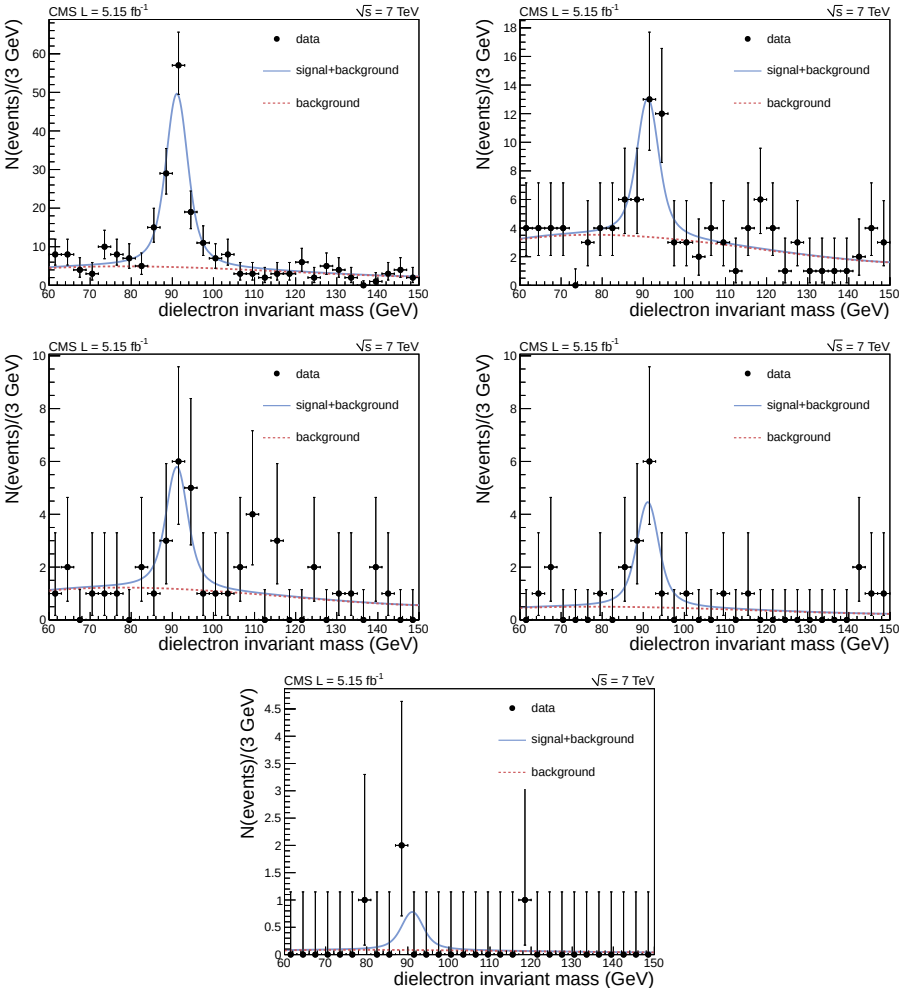

Figure A.17: Extended maximum likelihood fit of the dielectron invariant mass distribution for the signal extraction, in all bins of the differential cross section as a function of A_{ZBB}, with no cut on p_T^Z. From top left to bottom right: $A_{ZBB} < 0.2$, $0.2 < A_{ZBB} < 0.4$, $0.4 < A_{ZBB} < 0.6$, $0.6 < A_{ZBB} < 0.8$, $0.8 < A_{ZBB} < 1.0$.

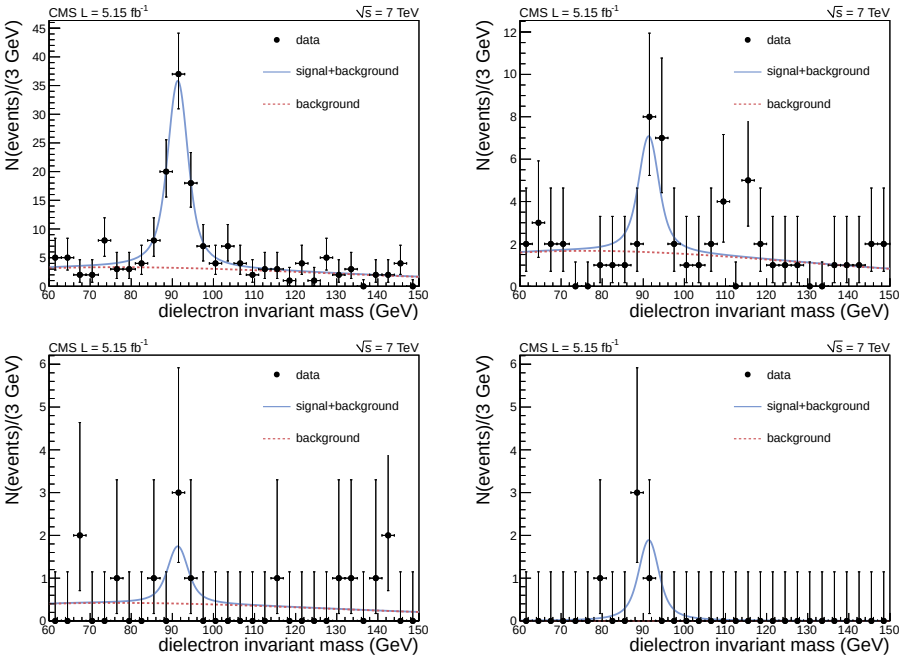

Figure A.18: Extended maximum likelihood fit of the dielectron invariant mass distribution for the signal extraction, in all bins of the differential cross section as a function of A_{ZBB}, for $p_T^Z > 50$ GeV. From top left to bottom right: $A_{ZBB} < 0.25$, $0.25 < A_{ZBB} < 0.5$, $0.5 < A_{ZBB} < 0.75$, $0.75 < A_{ZBB} < 1.0$.

A.2 Differential cross-section measurements

A.2.1 Differential cross section as a function of ΔR_{BB}

bin	dilepton channel	$N^{sig,\ell}$	$\mathcal{A}_\ell \times \epsilon_\ell$ ($\times 10^{-2}$)	$\frac{\mathcal{S}}{\epsilon_B}$ ($\times 10^{-2}$)	$\mathcal{P}$ ($\times 10^{-2}$)
0-0.7	$\mu\mu$	53.3±7.9±0.4	62.5±0.3	1045.3±83.3	85.7±3.2
	ee	28.5±5.8±0.2	51.4±5.1		
0.7-1.4	$\mu\mu$	38.1±7.3±0.6	64.4±0.3	1197.0±95.4	95.6±2.0
	ee	29.8±6.3±0.3	52.8±5.3		
1.4-2.1	$\mu\mu$	37.5±7.2±0.4	62.8±0.3	1300.2±103.7	95.2±1.7
	ee	26.2±6.5±0.5	54.7±5.5		
2.1-2.8	$\mu\mu$	45.6±8.2±0.6	64.3±0.3	1096.5±87.4	92.5±1.8
	ee	33.7±7.2±0.4	53.9±5.4		
2.8-3.5	$\mu\mu$	57.7±8.9±0.7	63.2±0.3	1042.0±83.1	92.8±1.7
	ee	35.9±7.2±0.5	52.4±5.2		
3.5-4.2	$\mu\mu$	19.3±4.8±0.1	63.5±0.3	1233.0±98.3	94.8±3.6
	ee	8.1±3.3±0.1	49.7±5.0		

Table A.1: Summary of the signal yields from the maximum likelihood fit, and of all the corrections for the cross-section measurement. Only statistical uncertainties are shown. No cut on p_T^Z is applied.

bin	dilepton channel	$N^{sig,\ell}$	$\mathcal{A}_\ell \times \epsilon_\ell$ ($\times 10^{-2}$)	$\frac{\mathcal{S}}{\epsilon_B}$ ($\times 10^{-2}$)	$\mathcal{P}$ ($\times 10^{-2}$)
0-0.84	$\mu\mu$	39.4±7.0±0.5	65.8±0.3	1061.6±74.1	86.2±3.5
	ee	29.3±5.7±0.2	55.2±5.5		
0.84-1.68	$\mu\mu$	37.6±6.9±0.4	65.8±0.3	968.6±67.7	97.0±1.6
	ee	18.0±5.1±0.3	55.7±5.6		
1.68-2.52	$\mu\mu$	32.7±6.8±0.4	64.4±0.3	1119.9±78.2	93.2±2.4
	ee	27.6±6.2±0.3	58.8±5.9		
2.52-3.36	$\mu\mu$	35.8±7.1±0.6	64.1±0.3	758.1±52.9	93.1±2.1
	ee	19.5±5.3±0.2	56.5±5.7		
3.36-4.2	$\mu\mu$	13.0±3.6±0.1	64.5±0.3	914.8±63.9	94.0±5.8
	ee	2.3±2.2±0.1	55.9±5.6		

Table A.2: Summary of the signal yields from the maximum likelihood fit, and of all the corrections for the cross-section measurement, for $p_T^Z > 50$ GeV. Only statistical uncertainties are shown.

A.2.2 Differential cross section as a function of $\Delta\Phi_{BB}$

bin	dilepton channel	$N^{sig,\ell}$	$\mathcal{A}_\ell \times \epsilon_\ell$ ($\times 10^{-2}$)	$\frac{\mathcal{S}}{\epsilon_B}$ ($\times 10^{-2}$)	$\mathcal{P}$ ($\times 10^{-2}$)
0-0.525	$\mu\mu$	65.4±8.9±0.5	63.1±0.3	1220.3±110.5	87.5±2.6
	ee	39.7±6.9±0.3	53.5±5.4		
0.525-1.05	$\mu\mu$	36.0±6.9±0.4	65.2±0.3	1101.8±99.8	96.2±1.7
	ee	29.1±6.3±0.4	57.5±5.8		
1.05-1.575	$\mu\mu$	28.2±6.2±0.3	64.1±0.3	1264.5±114.5	92.6±2.6
	ee	17.4±5.0±0.2	52.8±5.3		
1.575-2.1	$\mu\mu$	24.6±6.2±0.3	63.2±0.3	1225.2±110.9	95.6±1.9
	ee	24.9±6.3±0.4	53.7±5.4		
2.1-2.625	$\mu\mu$	33.6±7.1±0.4	65.3±0.3	1068.6±96.8	94.1±1.8
	ee	21.2±5.8±0.3	53.7±5.4		
2.625-3.15	$\mu\mu$	65.2±9.3±0.7	61.3±0.3	998.9±90.5	91.8±1.9
	ee	30.7±6.6±0.3	48.8±4.9		

Table A.3: Summary of the signal yields from the maximum likelihood fit, and of all the corrections for the cross-section measurement, with no cut on p_T^Z. Only statistical uncertainties are shown.

bin	dilepton channel	$N^{sig,\ell}$	$\mathcal{A}_\ell \times \epsilon_\ell$ ($\times 10^{-2}$)	$\frac{\mathcal{S}}{\epsilon_B}$ ($\times 10^{-2}$)	$\mathcal{P}$ ($\times 10^{-2}$)
0-0.63	$\mu\mu$	53.6±8.0±0.5	65.8±0.3	1151.6±83.8	87.4±2.9
	ee	40.0±6.8±0.3	57.8±5.8		
0.63-1.26	$\mu\mu$	35.4±6.7±0.4	67.2±0.3	933.4±67.9	96.6±1.7
	ee	20.6±5.2±0.3	58.6±5.9		
1.26-1.89	$\mu\mu$	24.5±5.8±0.3	63.7±0.3	975.2±71.0	95.3±2.3
	ee	12.8±4.5±0.3	55.1±5.5		
1.89-2.52	$\mu\mu$	14.2±5.0±0.3	64.6±0.3	861.3±62.7	93.7±2.6
	ee	18.0±5.0±0.1	56.1±5.6		
2.52-3.15	$\mu\mu$	31.7±6.3±0.3	61.6±0.3	804.6±58.6	91.9±2.9
	ee	7.8±3.8±0.2	53.2±5.3		

Table A.4: Summary of the signal yields from the maximum likelihood fit, and of all the corrections for the cross-section measurement, for $p_T^Z > 50$ GeV. Only statistical uncertainties are shown.

A.2.3 Differential cross section as a function of minΔR_{ZB}

bin	dilepton channel	$N^{sig,\ell}$	$\mathcal{A}_\ell \times \epsilon_\ell$ ($\times 10^{-2}$)	$\frac{\mathcal{S}}{\epsilon_B}$ ($\times 10^{-2}$)	$\mathcal{P}$ ($\times 10^{-2}$)
0-0.84	$\mu\mu$	13.5±4.9±0.3	62.8±0.3	1142.0±80.4	80.5±7.7
	ee	11.0±4.1±0.1	50.1±5.0		
0.84-1.68	$\mu\mu$	43.2±8.1±0.7	60.9±0.3	1160.6±81.7	89.4±3.4
	ee	28.8±7.1±0.6	52.8±5.3		
1.68-2.52	$\mu\mu$	79.7±10.5±1.0	61.2±0.3	1005.6±70.8	90.4±2.3
	ee	54.8±8.8±0.7	53.1±5.3		
2.52-3.36	$\mu\mu$	99.1±10.7±0.5	64.9±0.3	1109.1±78.1	96.0±1.2
	ee	55.3±8.4±0.5	53.2±5.3		
3.36-4.2	$\mu\mu$	13.1±4.4±0.2	64.5±0.3	1171.3±82.4	93.0±1.6
	ee	9.2±3.2±0.0	55.8±5.6		

Table A.5: Summary of the signal yields from the maximum likelihood fit, and of all the corrections for the cross-section measurement, with no cut on p_T^Z. Only statistical uncertainties are shown.

bin	dilepton channel	$N^{sig,\ell}$	$\mathcal{A}_\ell \times \epsilon_\ell$ ($\times 10^{-2}$)	$\frac{\mathcal{S}}{\epsilon_B}$ ($\times 10^{-2}$)	$\mathcal{P}$ ($\times 10^{-2}$)
0-1.05	$\mu\mu$	10.5±4.3±0.3	60.8±0.3	967.6±52.7	88.9±7.3
	ee	10.2±3.8±0.1	49.7±5.0		
1.05-2.1	$\mu\mu$	37.6±7.5±0.6	62.7±0.3	903.5±49.2	89.5±3.6
	ee	19.6±5.9±0.4	55.5±5.6		
2.1-3.15	$\mu\mu$	91.1±10.6±0.8	63.5±0.3	844.6±46.0	93.7±2.0
	ee	56.3±8.3±0.5	57.2±5.7		
3.15-4.2	$\mu\mu$	21.7±4.4±0.2	67.2±0.3	1046.9±57.0	93.0±1.8
	ee	12.7±3.9±0.1	58.9±5.9		

Table A.6: Summary of the signal yields from the maximum likelihood fit, and of all the corrections for the cross-section measurement, for $p_T^Z > 50$ GeV. Only statistical uncertainties are shown.

A.2.4 Differential cross section as a function of A_{ZBB}

bin	dilepton channel	$N^{sig,\ell}$	$\mathcal{A}_\ell \times \epsilon_\ell$ ($\times 10^{-2}$)	$\frac{\mathcal{S}}{\epsilon_B}$ ($\times 10^{-2}$)	$\mathcal{P}$ ($\times 10^{-2}$)
0-0.2	$\mu\mu$	167.1±14.7±1.6	64.1±0.3	1112.2±93.8	93.7±1.0
	ee	113.6±12.3±1.2	53.6±5.4		
0.2-0.4	$\mu\mu$	48.5±8.3±0.6	62.1±0.3	1120.2±94.5	91.7±2.0
	ee	24.5±6.7±0.5	51.2±5.1		
0.4-0.6	$\mu\mu$	24.3±5.9±0.3	62.0±0.3	1269.2±107.1	89.0±3.8
	ee	11.7±4.3±0.2	52.4±5.2		
0.6-0.8	$\mu\mu$	6.3±3.6±0.2	65.2±0.3	1182.3±99.8	99.3±0.7
	ee	10.1±3.5±0.1	51.9±5.2		
0.8-1	$\mu\mu$	4.6±2.4±0.0	61.9±0.3	1307.3±110.3	53.5±16.6
	ee	1.8±1.6±0.0	63.4±6.3		

Table A.7: Summary of the signal yields from the maximum likelihood fit, and of all the corrections for the cross-section measurement, with no cut on p_T^Z. Only statistical uncertainties are shown.

bin	dilepton channel	$N^{sig,\ell}$	$\mathcal{A}_\ell \times \epsilon_\ell$ ($\times 10^{-2}$)	$\frac{\mathcal{S}}{\epsilon_B}$ ($\times 10^{-2}$)	$\mathcal{P}$ ($\times 10^{-2}$)
0-0.25	$\mu\mu$	125.1±12.7±1.4	65.7±0.3	966.3±54.7	93.3±1.3
	ee	78.1±10.1±0.9	57.2±5.7		
0.25-0.5	$\mu\mu$	21.2±5.5±0.3	61.5±0.3	913.2±51.7	89.7±3.5
	ee	13.1±4.6±0.2	54.6±5.5		
0.5-0.75	$\mu\mu$	8.3±3.7±0.2	63.6±0.3	1080.5±61.1	100.0±0.0
	ee	3.2±2.2±0.0	52.0±5.2		
0.75-1	$\mu\mu$	4.8±2.4±0.0	65.1±0.3	965.9±54.7	68.9±18.0
	ee	4.5±2.0±0.0	58.3±5.8		

Table A.8: Summary of the signal yields from the maximum likelihood fit, and of all the corrections for the cross-section measurement, for $p_T^Z > 50$ GeV. Only statistical uncertainties are shown.

List of Figures

List of Tables

BIBLIOGRAPHY

[1] ATLAS Collaboration, "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC", *Phys. Lett.* **B716** (2012) 1, arXiv:1207.7214. doi:10.1016/j.physletb.2012.08.020. (Cited on pages 5 and 9.)

[2] CMS Collaboration, "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC", *Phys. Lett.* **B716** (2012) 30, arXiv:1207.7235. doi:10.1016/j.physletb.2012.08.021. (Cited on pages 5 and 9.)

[3] S. L. Glashow, "Partial-symmetries of weak interactions", *Nucl. Phys.* **22** (1961), no. 4, 579. doi:10.1016/0029-5582(61)90469-2. (Cited on page 7.)

[4] S. Weinberg, "A Model of Leptons", *Phys. Rev. Lett.* **19** (1967) 1264. doi:10.1103/PhysRevLett.19.1264. (Cited on page 7.)

[5] F. Englert and R. Brout, "Broken Symmetry and the Mass of Gauge Vector Mesons", *Phys. Rev. Lett.* **13** (1964) 321. doi:10.1103/PhysRevLett.13.321. (Cited on page 8.)

[6] P. W. Higgs, "Broken symmetries, massless particles and gauge fields", *Phys. Lett.* **12** (1964), no. 2, 132. doi:10.1016/0031-9163(64)91136-9. (Cited on page 8.)

[7] P. W. Higgs, "Broken Symmetries and the Masses of Gauge Bosons", *Phys. Rev. Lett.* **13** (1964) 508. doi:10.1103/PhysRevLett.13.508. (Cited on page 8.)

[8] LEP Working Group for Higgs boson searches, ALEPH, DELPHI, L3, OPAL Collaboration, "Search for the standard model Higgs boson at LEP", *Phys. Lett.* **B565** (2003) 61, arXiv:hep-ex/0306033. doi:10.1016/S0370-2693(03)00614-2. (Cited on page 8.)

[9] TEVNPH (Tevatron New Phenomena and Higgs Working Group), CDF, D0 Collaboration, "Combined CDF and D0 Search for Standard Model Higgs Boson Production with up to 10.0 fb^{-1} of Data", arXiv:1203.3774. (Cited on page 8.)

[10] CMS Collaboration, "Search for the standard model Higgs boson produced in association with W or Z bosons, and decaying to bottom quarks (LHCP 2013)", *CMS Physics Analysis Summary* **CMS PAS HIG-13-011** (2013). (Cited on page 10.)

[11] ATLAS Collaboration, "Measurement of the cross-section for b-jets produced in association with a Z boson at $\sqrt{s} = 7$ TeV with the ATLAS detector", *Phys. Lett.* **B706** (2012) 295, arXiv:1109.1403. doi:10.1016/j.physletb.2011.11.059. (Cited on page 10.)

[12] CMS Collaboration, "Measurement of the Z/gamma*+b-jet cross section in pp collisions at 7 TeV", *JHEP* **06** (2012) 126, arXiv:1204.1643. doi:10.1007/JHEP06(2012)126. (Cited on page 10.)

[13] W. J. S. Ellis, R. K. and B. R. Webber, "QCD and Collider Physics". Cambridge University Press, 1996. (Cited on page 10.)

[14] B. R. Webber, "Monte Carlo Simulation of Hard Hadronic Processes", *Annual Review of Nuclear and Particle Science* **36** (1986), no. 1, 253, arXiv:http://www.annualreviews.org/doi/pdf/10.1146/annurev.ns.36.120186.001345. doi:10.1146/annurev.ns.36.120186.001345. (Cited on page 14.)

[15] T. Sjostrand, S. Mrenna, and P. Z. Skands, "PYTHIA 6.4 Physics and Manual", *JHEP* **0605** (2006) 026, arXiv:hep-ph/0603175. doi:10.1088/1126-6708/2006/05/026. (Cited on page 15.)

[16] G. Corcella, I. Knowles, G. Marchesini et al., "HERWIG 6: An Event generator for hadron emission reactions with interfering gluons (including supersymmetric processes)", *JHEP* **0101** (2001) 010, arXiv:hep-ph/0011363. doi:10.1088/1126-6708/2001/01/010. (Cited on page 15.)

[17] S. Frixione, P. Nason, and C. Oleari, "Matching NLO computations with parton shower simulations: the POWHEG method", *JHEP* **2007** (2007), no. 11, 070. doi:10.1088/1126-6708/2007/11/070. (Cited on page 15.)

[18] J. Alwall, S. Hoche, F. Krauss et al., "Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions", *Eur. Phys. J.* **C53** (2008) 473, arXiv:0706.2569. doi:10.1140/epjc/s10052-007-0490-5. (Cited on page 16.)

[19] S. Catani, F. Krauss, R. Kuhn et al., "QCD matrix elements + parton showers", *JHEP* **0111** (2001) 063, arXiv:hep-ph/0109231. (Cited on pages 16 and 17.)

[20] M. Cacciari, G. P. Salam, and G. Soyez, "The anti-k_T jet clustering algorithm", *JHEP* **2008** (2008), no. 04, 063. (Cited on pages 16 and 61.)

[21] F. Maltoni, G. Ridolfi, and M. Ubiali, "b-initiated processes at the LHC: a reappraisal", *JHEP* **2012** (2012), no. 7, 1. doi:10.1007/JHEP07(2012)022. (Cited on page 17.)

[22] J. Alwall, P. Demin, S. de Visscher et al., "MadGraph/MadEvent v4: The New Web Generation", *JHEP* **0709** (2007) 028, arXiv:0706.2334. doi:10.1088/1126-6708/2007/09/028. (Cited on page 18.)

[23] J. Alwall, M. Herquet, F. Maltoni et al., "MadGraph 5 : Going Beyond", *JHEP* **1106** (2011) 128, arXiv:1106.0522. doi:10.1007/JHEP06(2011)128. (Cited on page 18.)

[24] M. L. Mangano, M. Moretti, F. Piccinini et al., "ALPGEN, a generator for hard multiparton processes in hadronic collisions", *JHEP* **0307** (2003) 001, arXiv:hep-ph/0206293. (Cited on page 18.)

[25] R. Frederix, S. Frixione, V. Hirschi et al., "W and $Z/\gamma*$ boson production in association with a bottom-antibottom pair", *JHEP* **1109** (2011) 61, arXiv:1106.6019. doi:10.1007/JHEP09(2011)061. (Cited on page 19.)

[26] S. Frixione, F. Stoeckli, P. Torrielli et al., "NLO QCD corrections in Herwig++ with MC@NLO", *JHEP* **1101** (2011) 53, arXiv:1010.0568. doi:10.1007/JHEP01(2011)053. (Cited on page 19.)

[27] G. Marchesini and B. Webber, "Simulation of QCD jets including soft gluon interference", *Nucl. Phys. B* **238** (1984), no. 1, 1. doi:http://dx.doi.org/10.1016/0550-3213(84)90463-2. (Cited on page 19.)

[28] B. Andersson, G. Gustafson, G. Ingelman et al., "Parton fragmentation and string dynamics", *Physics Reports* **97** (1983) 31. doi:http://dx.doi.org/10.1016/0370-1573(83)90080-7. (Cited on page 19.)

[29] T. Sjostrand and P. Z. Skands, "Multiple interactions and the structure of beam remnants", *JHEP* **0403** (2004) 053, arXiv:hep-ph/0402078. doi:10.1088/1126-6708/2004/03/053. (Cited on page 20.)

[30] O. S. Brüning, P. Collier, P. Lebrun et al., "LHC Design Report". CERN, Geneva, 2004. (Cited on page 23.)

[31] ATLAS Collaboration, "The ATLAS Experiment at the CERN Large Hadron Collider", *JINST* **3** (2008), no. 08, S08003. doi:10.1088/1748-0221/3/08/S08003. (Cited on page 23.)

[32] LHCb Collaboration, "The LHCb Detector at the LHC", *JINST* **3** (2008), no. 08, S08005. doi:10.1088/1748-0221/3/08/S08005. (Cited on page 23.)

[33] ALICE Collaboration, "The ALICE experiment at the CERN LHC", *JINST* **3** (2008), no. 08, S08002. doi:10.1088/1748-0221/3/08/S08002. (Cited on page 23.)

[34] "LEP design report", *CERN-LEP-84-01* (1984). Copies shelved as reports in LEP, PS and SPS libraries. (Cited on page 24.)

[35] CMS Collaboration, "CMS Physics Technical Design Report, Volume I: Detector Performance and Software", *CERN-LHCC-2006-001* **CMS-TDR-008-1** (2006). (Cited on page 26.)

[36] CMS Collaboration, "The CMS experiment at the CERN LHC", *JINST* **3** (2008) S08004. doi:10.1088/1748-0221/3/08/S08004. (Cited on page 26.)

[37] CMS Collaboration, "CMS Physics Technical Design Report, Volume II: Physics Performance", *Journal of Physics G: Nuclear and Particle Physics* **34** (2007), no. 6, 995. doi:10.1088/0954-3899/34/6/S01. (Cited on page 26.)

[38] CMS Collaboration, "The CMS magnet project: Technical Design Report", *CERN-LHCC-97-010* (1997). (Cited on page 26.)

[39] F. Kircher, P. Bredy, A. Calvo et al., "Final design of the CMS solenoid cold mass", *Applied Superconductivity, IEEE Transactions on* **10** (2000), no. 1, 407. doi:10.1109/77.828259. (Cited on page 26.)

[40] CMS Collaboration, "The CMS tracker system project: Technical Design Report", *CERN-LHCC-98-006* (1997). (Cited on page 28.)

[41] K. Arndt, G. Bolla, D. Bortoletto et al., "Silicon sensors development for the CMS pixel system", *Nucl. Instr. Meth. A* **511** (2003), no. 1-2, 106. doi:10.1016/S0168-9002(03)01773-X. (Cited on page 30.)

[42] H. Kaestli, M. Barbero, W. Erdmann et al., "Design and performance of the CMS pixel detector readout chip", *Nucl. Instr. Meth. A* **565** (2006), no. 1, 188. doi:10.1016/j.nima.2006.05.038. (Cited on page 30.)

[43] CMS Collaboration, "The CMS electromagnetic calorimeter project: Technical Design Report", *CERN-LHCC-97-033* (1997). (Cited on page 33.)

[44] CMS Collaboration, "CMS ECAL Preshower and Endcap Engineering Design Review v.2 - Preshower", *CERN-ECAL-EDR-4* (2000). (Cited on page 33.)

[45] CMS Collaboration, "The CMS hadron calorimeter project: Technical Design Report", *CERN-LHCC-97-031* (1997). (Cited on page 33.)

[46] CMS Collaboration, "The CMS muon project: Technical Design Report", *CERN-LHCC-97-032* (1997). (Cited on page 37.)

[47] CMS Collaboration, "CMS TriDAS project: Technical Design Report, Volume 1: The level-1 trigger", *CERN-LHCC-2000-038* (2000). (Cited on page 38.)

[48] CMS Collaboration, "CMS TriDAS project: Technical Design Report, Volume 2: Data acquisition and high-level trigger technical design report", *CERN-LHCC-2002-026* (2002). (Cited on pages 38 and 39.)

[49] CMS Collaboration, "The CMS high level trigger", *EPJ C - Particles and Fields* **46** (2006) 605. doi:10.1140/epjc/s2006-02495-8. (Cited on page 38.)

[50] CMS Collaboration, "A new technique for the reconstruction, validation, and simulation of hits in the CMS Pixel Detector", *CMS Physics Analysis Note* **CMS-AN-2007/033** (2007). (Cited on page 41.)

[51] CMS Collaboration, "A detailed simulation of the CMS pixel sensor", *CMS Physics Analysis Note* **CMS-AN-2002/027** (2002). (Cited on page 43.)

[52] E. Alagoz, V. Chiochia, and M. Swartz, "Simulation and hit reconstruction of irradiated pixel sensors for the CMS experiment", *Nucl. Instr. Meth.* **A566** (2006) 40. doi:10.1016/j.nima.2006.05.027. (Cited on page 43.)

[53] CMS Collaboration, "CMS Tracking Performance Results from early LHC Operation", *Eur. Phys. J.* **C70** (2010) 1165–1192, arXiv:1007.1988. doi:10.1140/epjc/s10052-010-1491-3. (Cited on pages 45 and 53.)

[54] CMS Collaboration, "Commissioning and performance of the CMS pixel tracker with cosmic ray muons", *JINST* **5** (2010), no. 03, T03007. doi:10.1088/1748-0221/5/03/T03007. (Cited on page 45.)

[55] CMS Collaboration, "Estimation of the Tracker Alignment Precision", *CMS Physics Analysis Note* **CMS-AN-12/235** (2011). (Cited on page 49.)

[56] CMS Collaboration, "Surface Shape of Silicon Tracker Modules: Parametrisation, Alignment and Hit Position Corrections", *CMS Physics Analysis Note* **CMS-AN-11/531** (2011). (Cited on page 49.)

[57] CMS Collaboration, "Description and performance of the CMS track and primary vertex reconstruction", *CMS Physics Analysis Note* **CMS-AN -11/172** (2011). (Cited on pages 53, 55, and 57.)

[58] R. E. Kalman, "A New Approach to Linear Filtering and Prediction Problems", *Transactions of the ASME, Journal of Basic Engineering* (1960), no. 82 (Series D), 35. (Cited on page 54.)

[59] CMS Collaboration, "Offline primary vertex reconstruction with Deterministic Annealing clustering", *CMS Internal Note* **CMS-IN -11/014** (2011). (Cited on page 57.)

[60] W. Waltenberger, R. Fruehwirth, and P. Vanlaer, "Adaptive vertex fitting", *J. Phys. G: Nuclear and Particle Physics* **34** (2007), no. 12, N343. doi:10.1088/0954-3899/34/12/N01. (Cited on page 57.)

[61] CMS Collaboration, "Performance of muon reconstruction and identification in pp collisions at $\sqrt{s} = 7$ TeV", *CMS Physics Analysis Summary* **CMS PAS MUO-10-004** (2010). (Cited on page 59.)

[62] CMS Collaboration, "Electron reconstruction and identification at $\sqrt{s} = 7$ TeV", *CMS Physics Analysis Summary* **CMS PAS EGM-10-004** (2010). (Cited on pages 60 and 78.)

[63] W. Adam, R. Fruehwirth, A. Strandlie et al., "Reconstruction of electrons with the Gaussian-sum filter in the CMS tracker at the LHC", *J. Phys. G: Nuclear and Particle Physics* **31** (2005), no. 9, N9. (Cited on page 60.)

[64] CMS Collaboration, "Particle-flow event reconstruction in CMS and performance for jets, taus and E_T^{miss}", *CMS Physics Analysis Summary* **CMS PAS PFT-09-001** (2009). (Cited on pages 60 and 77.)

[65] CMS Collaboration, "Identification of b-quark jets with the CMS experiment", *JINST* **8** (2013) P04013, arXiv:1211.4462. doi:10.1088/1748-0221/8/04/P04013. (Cited on pages 63, 69, and 149.)

[66] L. Wehrli, "Measurement of BB Angular Correlations based on Secondary Vertex Reconstruction in Proton-Proton collisions at $\sqrt{s} = 7$ TeV", *Ph.D. dissertation at ETH* (2011). (Cited on pages 65, 66, and 67.)

[67] J. Allison, K. Amako, J. Apostolakis et al., "Geant4 developments and applications", *Nuclear Science, IEEE Transactions on* **53** (2006), no. 1, 270. doi:10.1109/TNS.2006.869826. (Cited on page 80.)

[68] R. Castello et al., "Lepton efficiency and jets selection for the measurement of the $Z + b$ cross section", *CMS Analysis Note* **CERN-CMS-AN-11-264** (2011). (Cited on pages 81, 100, and 101.)

[69] CMS Collaboration, "Measurement of the $t\bar{t}$ production cross section in the all-jet final state in pp collisions at $\sqrt{s} = 7$ TeV", *JHEP* **1305** (2013) 065, arXiv:1302.0508. doi:10.1007/JHEP05(2013)065. (Cited on page 95.)

[70] CMS Collaboration, "Measurement of the $t\bar{t}$ production cross section in pp collisions at $\sqrt{s} = 7$ TeV with lepton + jets final states", *Phys. Lett.* **B720** (2013) 83–104, arXiv:1212.6682. doi:10.1016/j.physletb.2013.02.021. (Cited on page 95.)

[71] L. Lyons, D. Gibaut, and P. Clifford, "How to combine correlated estimates of a single physical quantity", *Nucl. Instr. Meth. A* **270** (1988) 110. doi:10.1016/0168-9002(88)90018-6. (Cited on page 102.)

[72] A. Valassi, "Combining correlated measurements of several different physical quantities", *Nucl. Instr. Meth. A* **500** (2003) 391. doi:10.1016/S0168-9002(03)00329-2. (Cited on page 102.)

[73] CMS Collaboration, "Absolute Calibration of the Luminosity Measurement at CMS: Winter 2012 Update", *CMS Physics Analysis Summary* **CMS-PAS-SMP-12-008** (2012). (Cited on page 112.)

[74] R. Gavin, Y. Li, F. Petriello et al., "FEWZ 2.0: A code for hadronic Z production at next-to-next-to-leading order", *Comput. Phys. Commun.* **182** (2011) 2388–2403, arXiv:1011.3540. doi:10.1016/j.cpc.2011.06.008. (Cited on page 114.)